

ON THE SPECTRAL DENSITY OF THE WAVELET COEFFICIENTS OF LONG MEMORY TIME SERIES WITH APPLICATION TO THE LOG-REGRESSION ESTIMATION OF THE MEMORY PARAMETER

E. MOULINES, F. ROUEFF, AND M.S. TAQQU

CNRS LTCI and Boston University

ABSTRACT. In recent years, methods to estimate the memory parameter using wavelet analysis have gained popularity in many areas of science. Despite its widespread use, a rigorous semi-parametric asymptotic theory, comparable to the one developed for Fourier methods, is still missing. In this contribution, we adapt to the wavelet setting the classical semi-parametric framework introduced by Robinson and his co-authors for estimating the memory parameter of a (possibly) non-stationary process. Our results apply to a class of wavelets with bounded supports, which include but are not limited to Daubechies wavelets. We derive an explicit expression of the spectral density of the wavelet coefficients and show that it can be approximated, at large scales, by the spectral density of the continuous-time wavelet coefficients of fractional Brownian motion. We derive an explicit bound for the difference between the spectral densities. As an application, we obtain minimax upper bounds for the log-scale regression estimator of the memory parameter for a Gaussian process and we derive an explicit expression of its asymptotic variance.

Date: June 6, 2006.

1991 Mathematics Subject Classification. Primary 62M10, 62M15, 62G05 Secondary: 60G18.

Key words and phrases. Wavelet analysis, long range dependence, semi-parametric estimation.

Murad S. Taqqu would like to thank l’École Normale Supérieure des Télécommunications in Paris for their hospitality. This research was partially supported by the NSF Grant DMS-0505747 at Boston University.

Corresponding author: F. Roueff, GET/Télécom Paris, CNRS LTCI, 46, rue Barrault, 75634 Paris Cédex 13, France.

1. INTRODUCTION

Let $X \stackrel{\text{def}}{=} \{X_k\}_{k \in \mathbb{Z}}$ be a real-valued process, not necessarily stationary and let $\Delta^K X$ denote its K -th order difference. The first order difference is $[\Delta X]_k \stackrel{\text{def}}{=} X_k - X_{k-1}$ and Δ^K is defined recursively. The process X is said to have *memory parameter* d , $d \in \mathbb{R}$ (in short, is an $M(d)$ process) if for any integer $K > d - 1/2$, the K -th order difference process $\Delta^K X$ is weakly stationary with spectral density function

$$f_{\Delta^K X}(\lambda) \stackrel{\text{def}}{=} |1 - e^{-i\lambda}|^{2(K-d)} f^*(\lambda) \quad \lambda \in (-\pi, \pi), \quad (1)$$

where f^* is a non-negative symmetric function which is bounded on $(-\pi, \pi)$ and is bounded away from zero in a neighborhood of the origin. $M(d)$ processes encompass both stationary and non-stationary processes, depending on the value of the memory parameter d . When $d < 1/2$, the process X is covariance stationary and its spectral density is given by

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda). \quad (2)$$

The process X is said to have long-memory if $0 < d < 1/2$, short-memory if $d = 0$ and negative memory if $d < 0$; the process is not invertible if $d < -1/2$. When $d > 1/2$, the process is non stationary. In this case, the f in (2) is not integrable on $[-\pi, \pi]$ and is therefore not a spectral density. In the terminology of Yaglom (1958), this f is called a *generalized spectral density*. It corresponds to a process X whose increments of sufficiently high order are covariance stationary.

The memory parameter d plays a central role in the definition of $M(d)$ processes and is often the focus of empirical interest. In the parametric case one can use approximate MLE methods (Fox and Taqqu (1986)) or MLE (Dahlhaus (1989)). In the semi-parametric case (2) where only a class of functions f^* is specified, two types of methods have emerged to estimate the memory parameter d : Fourier and Wavelet methods. Frequency-domain techniques are now well documented and understood (see for instance Hurvich and Ray (1995), Velasco (1999), Velasco and Robinson (2000) and Hurvich et al. (2002)).

In this paper, we focus on wavelet methods and consider the regression estimator introduced in Abry and Veitch (1998), which involves estimating d using the slope of the regression of the logarithm of the scale spectrum on the scale index. This estimator is now widely used in many different fields (see *e.g.* Veitch and Abry (1999) for applications to network traffic; Percival and Walden (2000) and Papanicolaou and Sølna (2003) for applications in physical sciences; see *e.g.* Gençay et al. (2002) and Bayraktar et al. (2004) for

applications in finance). The regression estimator is well-suited to process large data sets, since it has low computational complexity due to the pyramidal algorithm for computing the details coefficients. Also, it is robust with respect to additive polynomial trends (see for instance Veitch and Abry (1999) and Craigmile et al. (2005)). In Moulines, Roueff and Taqqu (2005), we study another estimator of d obtained by adapting the local Whittle estimator to the wavelet context.

Despite its widespread use, a rigorous semi-parametric asymptotic theory of the regression estimator, comparable to the one developed for corresponding estimators based on the periodogram, is still missing (the concluding remarks in Velasco (1999) about “the lack of rigorous asymptotic theory (...) if the spectral density is not proportional to λ^{-2d} for all frequencies” for wavelet-based estimates are still valid). There are results in a related parametric framework (see Bardet (2002) and Bayraktar et al. (2004)). To the best of our knowledge, the only attempt in a semi-parametric setting is due to Bardet et al. (2000). The process, however, is supposed to be observed in continuous-time – discretization issues were not discussed – and the results do not directly translate to discrete-time observations in a semi-parametric framework. The main objective of this paper is to fill this gap.

The paper is organized as follows. Examples of $M(d)$ processes are given in Section 2. In Section 3, we introduce wavelets and wavelet transforms for time-series. We do not assume that the wavelets are orthonormal nor that they result from a multiresolution analysis. In Section 4, we derive an explicit expression for the covariance and spectral density of the wavelet coefficients of an $M(d)$ process at a given scale. We extend this result to pairs of scales by grouping, in an appropriate way, the wavelet coefficients. The results apply to a general class of wavelets with bounded supports, which include but are not limited to Daubechies wavelets. If f^* belongs to a class of smooth functions $\mathcal{H}(\beta, L)$ with smoothness exponent β defined in (22), we show that the spectral density of the wavelet coefficients of an $M(d)$ process can be approximated, at large scales, by the spectral density of the wavelet coefficients of fractional Brownian motion (FBM) and derive an explicit bound for the difference between these two quantities. Our result holds not only for $d \in (1/2, 3/2)$, which corresponds to the standard range for the Hurst index, $H = d - 1/2 \in (0, 1)$, but for all $d \in \mathbb{R}$ by interpreting the corresponding FBM as a generalized process with spectral density $|\lambda|^{-2d}$, $\lambda \in \mathbb{R}$. We show that the relative L^∞ error between the spectral densities of the wavelet coefficients decreases exponentially fast to zero with a rate given by the smoothness exponent β of f^* . In Section 5, we consider (possibly non-stationary) Gaussian

processes and obtain an explicit expression for the limiting variance of the estimator of d based on the regression of the log-scale spectrum. We show that this estimator is rate optimal in the minimax sense. Sections 6 and 7 contain proofs. Appendix A involves approximations of wavelet filter transfer functions. We derive in Appendix B an inequality for the mean and the covariance of the logarithm of quadratic forms of Gaussian variables.

2. EXAMPLES

Stationarity of the increments is commonly assumed in time-series analysis. In ARIMA models, for example, (1) holds with $d = K$ integer and with f^* equal to the spectral density of an autoregressive moving average short-memory process. If $d \in \mathbb{R}$ and $f^* \equiv \sigma^2$ in (2), one gets the so-called fractionally integrated white noise process, ARFIMA(0,d,0). The choice $d \in \mathbb{R}$ and

$$f_{\text{ARMA}}^*(\lambda) = \sigma^2 \frac{\left|1 - \sum_{k=1}^q \theta_k e^{-i\lambda k}\right|^2}{\left|1 - \sum_{k=1}^p \phi_k e^{-i\lambda k}\right|^2}, \quad \lambda \in (-\pi, \pi), \quad (3)$$

with $1 - \sum_{k=1}^p \phi_k z^k \neq 0$ for $|z| = 1$ and $1 - \sum_{k=1}^q \theta_k \neq 0$ (so that $f_{\text{ARMA}}^*(0) \neq 0$) leads to the class of ARFIMA(p, d, q) processes.

Another example is $\{B_H(k)\}_{k \in \mathbb{Z}}$, a discrete-time version of fractional Brownian motion (FBM) $\{B_H(t), t \in \mathbb{R}\}$ with Hurst index $H \in (0, 1)$. The latter is a centered Gaussian process with covariance

$$R_H(t, s) \stackrel{\text{def}}{=} \mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2} \{ |t|^{2H} + |s|^{2H} - |t-s|^{2H} \}.$$

The process $\{B_H(k)\}_{k \in \mathbb{Z}}$ is increment stationary ($K = 1$) and its generalized spectral density is given up to a multiplicative constant (see Samorodnitsky and Taqqu (1994)) by

$$f_{\text{FBM}}(\lambda) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{-2H-1}, \quad \lambda \in (-\pi, \pi).$$

We can express it in the form (2),

$$f_{\text{FBM}}(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_{\text{FBM}}^*(\lambda), \quad (4)$$

by setting $d = H + 1/2 \in (1/2, 3/2)$ and

$$f_{\text{FBM}}^*(\lambda) = \left| \frac{2 \sin(\lambda/2)}{\lambda} \right|^{2H+1} + |2 \sin(\lambda/2)|^{2H+1} \sum_{k \neq 0} |\lambda + 2k\pi|^{-2H-1}. \quad (5)$$

Observe that $f_{\text{FBM}}^*(0) = 1$ and that it is bounded on $(-\pi, \pi)$.

The process $G_H = \Delta B_H$ is fractional Gaussian noise (FGN). It is a stationary Gaussian process with spectral density proportional to (4), but with $d = H - 1/2 \in (-1/2, 1/2)$.

By convention, throughout the paper, while d may take values in \mathbb{R} , H will be always restricted to take values in $(0, 1)$.

3. DISCRETE WAVELET TRANSFORM

In this section, we introduce the main concepts required to define a discrete wavelet transform. Denote by $L^2(\mathbb{R})$ the set of square integrable functions with respect to the Lebesgue measure. Let $\phi \in L^2(\mathbb{R})$ and $\psi \in L^2(\mathbb{R})$ be two functions and define their Fourier transforms as

$$\widehat{\phi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \phi(t) e^{-i\xi t} dt \quad \text{and} \quad \widehat{\psi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \psi(t) e^{-i\xi t} dt.$$

Consider the following assumptions:

- (W-1) ϕ and ψ are compactly-supported, integrable, and $\widehat{\phi}(0) = \int_{-\infty}^{\infty} \phi(t) dt = 1$ and $\int_{-\infty}^{\infty} \psi^2(t) dt = 1$.
- (W-2) There exists $\alpha > 1$ such that $\sup_{\xi \in \mathbb{R}} |\widehat{\psi}(\xi)| (1 + |\xi|)^{\alpha} < \infty$.
- (W-3) The function ψ has M vanishing moments, *i.e.* $\int_{-\infty}^{\infty} t^m \psi(t) dt = 0$ for all $m = 0, \dots, M-1$
- (W-4) The function $\sum_{k \in \mathbb{Z}} k^m \phi(-k)$ is a polynomial of degree m for all $m = 0, \dots, M-1$.

Assumption (W-1) implies that $\widehat{\phi}$ and $\widehat{\psi}$ are everywhere infinitely differentiable. The exponent α in (W-2) is related to the rate of decrease of the Fourier transform $\widehat{\psi}$ of the wavelet ψ and hence to the regularity of ψ . Under (W-1), (W-3) is equivalent to asserting that the first $M-1$ derivatives of $\widehat{\psi}$ vanish at the origin. This implies, using a Taylor expansion, that

$$|\widehat{\psi}(\xi)| = O(|\xi|^M) \quad \text{as} \quad \xi \rightarrow 0. \quad (6)$$

By (Cohen, 2003, Theorem 2.8.1, Page 90), under (W-1), (W-4) is equivalent to

$$\sup_{k \neq 0} |\widehat{\phi}(\xi + 2k\pi)| = O(|\xi|^M) \quad \text{as} \quad \xi \rightarrow 0. \quad (7)$$

Adopting the engineering convention that large values of the scale index j correspond to coarse scales (low frequencies), we define the family $\{\psi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ of translated and dilated functions

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k). \quad (8)$$

Many authors suppose that the $\psi_{j,k}$ are orthogonal and even that they are generated by a multiresolution analysis (MRA). Assumptions (W-1)-(W-4) are indeed quite standard in the context of a multiresolution analysis (in which case, ϕ is the scaling function and ψ is the associated wavelet), see for instance Cohen (2003). In this paper, we do not assume that wavelets are orthonormal nor that they are associated to a multiresolution analysis. We may therefore work with other convenient choices for ϕ and ψ as long as (W-1)-(W-4) are satisfied. A simple example is to set, for some positive integer N ,

$$\phi(x) \stackrel{\text{def}}{=} \mathbb{1}_{[0,1]}^{\otimes N}(x) \quad \text{and} \quad \psi(x) \stackrel{\text{def}}{=} \frac{d^N}{dx^N} \mathbb{1}_{[0,1]}^{\otimes 2N}(x),$$

where $\mathbb{1}_A$ is the indicator function of the set A and for a non negative function f , $f^{\otimes N}$ denotes the N -th self-convolution of f . It follows that

$$|\widehat{\phi}(\xi)| = |2 \sin(\xi/2)/\xi|^N \quad \text{and} \quad \widehat{\psi}(\xi) = |\xi|^N |2 \sin(\xi/2)/\xi|^{2N}.$$

Using (6) and (7), one easily checks that (W-1)-(W-4) are satisfied with M and α equal to N . Of course the family of functions $\{\psi_{j,k}\}$ are not orthonormal for this choice of the wavelet function ψ (and the function ϕ is not associated to a MRA). Nevertheless, to ease references to previously reported works, with a slight abuse in the terminology, we still call ϕ and ψ the scaling and the wavelet functions.

Having defined the functions ϕ and ψ , we now define what we call the Discrete Wavelet Transform in discrete time. Start with a real-valued sequence $\{x_k, k \in \mathbb{Z}\}$. Using the scaling function ϕ , we first associate to the sequence $\{x_k, k \in \mathbb{Z}\}$ the continuous-time functions

$$\mathbf{x}_n(t) \stackrel{\text{def}}{=} \sum_{k=1}^n x_k \phi(t - k) \quad \text{and} \quad \mathbf{x}(t) \stackrel{\text{def}}{=} \sum_{k \in \mathbb{Z}} x_k \phi(t - k), \quad t \in \mathbb{R}. \quad (9)$$

The wavelet coefficients involve $\{\mathbf{x}(t), t \in \mathbb{R}\}$ and are defined as

$$W_{j,k}^{\mathbf{x}} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathbf{x}(t) \psi_{j,k}(t) dt \quad j \geq 0, k \in \mathbb{Z}. \quad (10)$$

Without loss of generality we may suppose that the support of the scaling function ϕ is included in $(-T, 0)$ for some integer $T \geq 1$. Then $\mathbf{x}_n(t) = \mathbf{x}(t)$ for all $t \in (0, n - T + 1)$. We may also suppose that the support of the wavelet function ψ is included in $(0, T)$. Then, the support of $\psi_{j,k}$ is included in the interval $(2^j k, 2^j (k + T))$. Hence

$$W_{j,k}^{\mathbf{x}} = W_{j,k}^{\mathbf{x}_n} = \int_{-\infty}^{\infty} \mathbf{x}_n(t) \psi_{j,k}(t) dt, \quad (11)$$

when $(2^j k, 2^j(k+T)) \subseteq (0, n-T+1)$, that is, for all $(j, k) \in \mathcal{I}_n$, where

$$\mathcal{I}_n \stackrel{\text{def}}{=} \{(j, k) : j \geq 0, 0 \leq k \leq 2^{-j}(n-T+1) - T\}. \quad (12)$$

For any j , the wavelet coefficients $\{W_{j,k}^{\mathbf{x}}\}_{k \in \mathbb{Z}}$ are obtained by discrete convolution and downsampling. More precisely, under (W-1), for all $j \geq 0, k \in \mathbb{Z}$, (9) and (10) imply

$$W_{j,k}^{\mathbf{x}} = \sum_{l \in \mathbb{Z}} x_l h_{j,2^j k - l} = (h_{j,\cdot} \star \mathbf{x})_{2^j k} = (\downarrow^j [h_{j,\cdot} \star \mathbf{x}])_k, \quad (13)$$

where \star denotes the convolution of discrete sequences,

$$h_{j,l} \stackrel{\text{def}}{=} 2^{-j/2} \int_{-\infty}^{\infty} \phi(t+l) \psi(2^{-j}t) dt,$$

and, for any sequence $\{c_k\}_{k \in \mathbb{Z}}$ and any integer l , \downarrow^j is the downsampling operator defined as $(\downarrow^j c)_k = c_{2^j k}$. Define, for all $j \geq 0$, the discrete Fourier transform of $\{h_{j,l}\}_{l \in \mathbb{Z}}$ as

$$H_j(\lambda) \stackrel{\text{def}}{=} \sum_{l \in \mathbb{Z}} h_{j,l} e^{-i\lambda l} = 2^{-j/2} \int_{-\infty}^{\infty} \sum_{l \in \mathbb{Z}} \phi(t+l) e^{-i\lambda l} \psi(2^{-j}t) dt. \quad (14)$$

Since ϕ and ψ have compact support, the sum in (14) has a finite number of non-vanishing terms and H_j is a trigonometric polynomial.

Remark 1. By Corollary 10, there exists an integer j_0 only depending on ϕ and ψ such that, for all $j \geq j_0$, the trigonometric polynomial $H_j(\lambda)$ is not identically zero. In the case of a multiresolution analysis, the father and mother wavelets are defined in such a way that $j_0 = 0$. In the general case, by dilating ψ appropriately, or, in other words, by changing the reference scale, one can impose $j_0 = 0$, which is assumed in the sequel.

Under assumption (W-4), $t \mapsto \sum_{l \in \mathbb{Z}} \phi(t+l) l^m$ is a polynomial of degree m and (W-3) therefore implies that, for all $j \geq 0$ and all $m = 0, \dots, M-1$,

$$\sum_{l \in \mathbb{Z}} h_{j,l} l^m = 2^{-j/2} \int_{-\infty}^{\infty} \psi(2^{-j}t) \sum_{l \in \mathbb{Z}} \phi(t+l) l^m dt = 0. \quad (15)$$

Now consider $P_j(x) = \sum_{l \in \mathbb{Z}} h_{j,l} x^l$ and observe that (15) implies $P_j(1) = 0, P'_j(1) = 0, \dots, P_j^{(M-1)}(1) = 0$, and hence $H_j(\lambda) = P_j(e^{-i\lambda})$ factors as

$$H_j(\lambda) = (1 - e^{-i\lambda})^M \tilde{H}_j(\lambda), \quad (16)$$

where $\tilde{H}_j(\lambda)$ is also a trigonometric polynomial. The wavelet coefficient (13) may therefore be computed as

$$W_{j,k}^{\mathbf{x}} = (\downarrow^j [\tilde{h}_{j,\cdot} \star \Delta^M \mathbf{x}])_k \quad (17)$$

where $\{\tilde{h}_{j,l}\}_{l \in \mathbb{Z}}$ are the coefficients of the trigonometric polynomial \tilde{H}_j and $\Delta^M \mathbf{x}$ is the M -th order difference of the sequence \mathbf{x} . In other words, the use of a wavelet and a scaling function satisfying (W-4) and (W-3) implicitly perform a M -th order differentiation of the time-series. Therefore, we may work with a K -th order integrated processes X without specific preprocessing, provided that $M \geq K$. This is in sharp contrast with Fourier methods, where the time series must be explicitly differentiated at least K times and a data taper must be applied on the differenced series to avoid frequency-domain leakage (see, for instance, Hurvich et al. (2002)).

4. SPECTRAL DENSITY OF THE WAVELET COEFFICIENTS

Because the wavelet coefficients at a given scale are obtained by applying time-invariant linear filters, computing the covariance of the wavelet coefficients of K -th order stationary processes is an easy exercise. The following proposition provides an integral expression for the covariance between two wavelet coefficients on possibly different scales, expressed in terms of the transfer function H_j of the wavelet filters and the generalized spectral density of the process X . This proposition extends Theorem 2 in Masry (1993) on the spectral measure of the DWT coefficients of increment stationary continuous-time processes to the discrete-time setting and Lemma 1 in Craigmile and Percival (2005) to functions ψ and ϕ that do not necessarily define a multiresolution analysis.

Proposition 1. *Let X be a K -th order integrated process with generalized spectral density f . Assume (W-1)-(W-4) with $M \geq K$. Then, for all $j, j' \geq 0$ and $k, k' \in \mathbb{Z}$,*

$$\text{Cov}(W_{j,k}^X, W_{j',k'}^X) = \int_{-\pi}^{\pi} e^{i\lambda(k2^j - k'2^{j'})} f(\lambda) H_j(\lambda) \overline{H_{j'}(\lambda)} d\lambda, \quad (18)$$

where the wavelet coefficient $W_{j,k}^X$ is defined in (10).

The proof follows from elementary results on time-invariant linear filtering of covariance stationary processes, using (16), (17), applied to $\Delta^M X$, which is covariance stationary with spectral density $|1 - e^{i\lambda}|^{2M} f(\lambda)$.

By (17), for a given scale j , the process $\{W_{j,k}^X\}_{k \in \mathbb{Z}}$ is covariance stationary. The situation is more complicated when considering two different scales $j \neq j'$, because the two-dimensional sequence $\{[W_{j,k}^X, W_{j',k}^X]^T\}_{k \in \mathbb{Z}}$, with T denoting the transpose, is not stationary for $j \neq j'$. This is a consequence of the pyramidal wavelet scheme, where at scale j , the

wavelet coefficients are downsampled by a factor 2^j which depends on j since $\psi_{j,k}$ in (8) can be expressed as

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}(t - 2^j k)) .$$

Thus, to obtain a stationary sequence, one should consider the process $\{[W_{j,k}^X, W_{j',2^{j-j'}k}^X]^T\}_{k \in \mathbb{Z}}$ for $j > j'$, which involves a downsampled subsequence of the coefficients at the finer scale j' . One can also consider the process $\{[W_{j,k}^X, W_{j',2^{j-j'}k+v}^X]^T\}_{k \in \mathbb{Z}}$ for $j > j'$, which includes a translation of the location index of the second component by v . It turns out that the most convenient is to merge the processes corresponding to $v = 0, \dots, 2^{j-j'} - 1$ and hence to consider the *between-scale* process

$$\{[W_{j,k}^X, \mathbf{W}_{j,k}^X(j - j')^T]^T\}_{k \in \mathbb{Z}} , \quad (19)$$

where for any $u = 0, 1, \dots, j$,

$$\mathbf{W}_{j,k}^X(u) \stackrel{\text{def}}{=} [W_{j-u,2^u k}^X, W_{j-u,2^u k+1}^X, \dots, W_{j-u,2^u k+2^u-1}^X]^T \quad (20)$$

is a 2^u -dimensional vector of wavelet coefficients at scale $j' = j - u$. The vector $\mathbf{W}_{j,k}^X(u)$ involves all possible translations of the position index $2^u k$ by $v = 0, 1, \dots, 2^u - 1$. The index u in (20) denotes the scale difference $j - j' \geq 0$ between the finest scale j' and the coarsest scale j . Observe that $\mathbf{W}_{j,k}^X(0)$ ($u = 0$) is the scalar $W_{j,k}^X$.

One should view the between-scale process (19) as a pair made up of the scalar process $\{W_{j,k}^X\}_{k \in \mathbb{Z}}$ and the vector process $\{\mathbf{W}_{j,k}^X(j - j')\}_{k \in \mathbb{Z}}$. We shall now express their cross spectral density in terms of the generalized spectral density of X and the transfer function of the wavelet filters folded on the interval $[-\pi, \pi]$. By setting $j' = j$ or equivalently $u = 0$ we obtain the spectral density of the “within scale” process $\{W_{j,k}^X\}_{k \in \mathbb{Z}}$.

Corollary 2. Define for all $0 \leq u \leq j$ and $\lambda \in (-\pi, \pi)$,

$$\mathbf{D}_{j,u}(\lambda; f, \phi, \psi) \stackrel{\text{def}}{=} \sum_{l=0}^{2^j-1} \mathbf{e}_u(\lambda + 2l\pi) f(2^{-j}(\lambda + 2l\pi)) 2^{-j/2} H_j(2^{-j}(\lambda + 2l\pi)) 2^{-(j-u)/2} \overline{H_{j-u}(2^{-j}(\lambda + 2l\pi))} , \quad (21)$$

where for all $\xi \in \mathbb{R}$,

$$\mathbf{e}_u(\xi) \stackrel{\text{def}}{=} 2^{-u/2} [1, e^{-i2^{-u}\xi}, \dots, e^{-i(2^u-1)2^{-u}\xi}]^T .$$

Then

$$\text{Cov}_f(W_{j,k}^X, \mathbf{W}_{j,k'}^X(u)) = \int_{-\pi}^{\pi} e^{i\lambda(k-k')} \mathbf{D}_{j,u}(\lambda; f, \phi, \psi) d\lambda .$$

In other words,

- for all $j \geq 0$, the within-scale process $\{W_{j,k}^X\}_{k \in \mathbb{Z}}$ is covariance stationary with spectral density $\mathbf{D}_{j,0}(\cdot; f, \phi, \psi)$,
- for all $j \geq u > 0$, the between-scale process $\{[W_{j,k}^X, \mathbf{W}_{j,k}^X(u)^T]^T\}_{k \in \mathbb{Z}}$ is covariance stationary with cross spectral density $\mathbf{D}_{j,u}(\cdot; f, \phi, \psi)$.

Note that $\mathbf{D}_{j,u}$ is a 2^u -dimensional vector and, in particular, $\mathbf{D}_{j,0}$ is a scalar. The 2^u -dimensional vector $\mathbf{e}_u(\xi)$ has Euclidean norm $|\mathbf{e}_u(\xi)| = 1$.

Proof. Let $j \geq u \geq 0$. By (18), we have, for all $k, k' \in \mathbb{Z}$ and $v = 0, \dots, 2^u - 1$,

$$\text{Cov}_f(W_{j,k}^X, W_{j-u, k' 2^u + v}^X) = \int_0^{2\pi} e^{i\lambda(k2^j - (k'2^u + v)2^{j-u})} f(\lambda) H_j(\lambda) \overline{H_{j-u}(\lambda)} d\lambda.$$

The exponential can be factorized as $e^{i2^j \lambda(k-k')} e^{-iv2^{-u}2^j \lambda}$. Hence

$$\begin{aligned} \text{Cov}_f(W_{j,k}^X, \mathbf{W}_{j,k'}^X(u)) &= \int_0^{2\pi} e^{i\lambda 2^j (k-k')} \mathbf{e}_u(2^j \lambda) f(\lambda) H_j(\lambda) 2^{u/2} \overline{H_{j-u}(\lambda)} d\lambda \\ &= \int_0^{2^{j+1}\pi} e^{i\lambda(k-k')} \mathbf{e}_u(\lambda) f(2^{-j}\lambda) 2^{-j/2} H_j(2^{-j}\lambda) 2^{-(j-u)/2} \overline{H_{j-u}(2^{-j}\lambda)} d\lambda. \end{aligned}$$

The result is obtained by *folding* and shifting the previous integral as follows, setting $g(\lambda) = \mathbf{e}_u(\lambda) f(2^{-j}\lambda) 2^{-j/2} H_j(2^{-j}\lambda) 2^{-(j-u)/2} \overline{H_{j-u}(2^{-j}\lambda)}$,

$$\begin{aligned} \int_0^{2^{j+1}\pi} e^{i\lambda(k-k')} g(\lambda) d\lambda &= \sum_{l=0}^{2^j-1} \int_{2l\pi}^{2(l+1)\pi} e^{i\lambda(k-k')} g(\lambda) d\lambda \\ &= \sum_{l=0}^{2^j-1} \int_0^{2\pi} e^{i\lambda(k-k')} g(\lambda + 2l\pi) d\lambda = \int_0^{2\pi} e^{i\lambda(k-k')} \left(\sum_{l=0}^{2^j-1} g(\lambda + 2l\pi) \right) d\lambda. \end{aligned}$$

The function in parentheses is (2π) -periodic because $\sum_{l=0}^{2^j-1} g(\lambda + 2(l+1)\pi) = \sum_{l=1}^{2^j-1} g(\lambda + 2l\pi) + g(\lambda + 2^j(2\pi)) = \sum_{l=0}^{2^j-1} g(\lambda + 2l\pi)$ since g is $2^j(2\pi)$ -periodic. Hence $\int_0^{2\pi}$ can be replaced by $\int_{-\pi}^{\pi}$, which gives the result. \square

We now apply the preceding results to the class of processes with memory parameter $d \in \mathbb{R}$ (see (1)). To obtain error bounds on the variance and the spectral density of the wavelet coefficients, some additional assumptions are required on the smoothness of f^* at zero frequency.

Definition 1. For $0 < \beta \leq 2$ and $L > 0$, define the function class $\mathcal{H}(\beta, L)$ as the set of even non-negative functions g on $[-\pi, \pi]$ such that, for all $\lambda \in [-\pi, \pi]$,

$$|g(\lambda) - g(0)| \leq L g(0) |\lambda|^\beta. \quad (22)$$

This type of assumption is typical in the *semi-parametric* estimation setting (see for instance Robinson (1995) and Moulines and Soulier (2002)). The larger the value of β , the smoother the function at the origin. For g even – as assumed – and infinitely differentiable, $g'(0) = 0$ and hence, by a Taylor expansion, (22) holds with $\beta = 2$.

Since, for instance, f_{ARMA}^* in (3) is infinitely differentiable, it belongs to $\mathcal{H}(2, L)$ for some L . As for FBM and FGN, consider f_{FBM}^* in (5). The first term in the RHS is $1 + O(\lambda^2)$ and the second is $O(|\lambda|^{2H+1})$. Hence, for some positive constant L , $|f_{\text{FBM}}^*(\lambda) - f_{\text{FBM}}^*(0)| \leq L f_{\text{FBM}}^*(0) |\lambda|^{(2H+1)\wedge 2}$, where $a \wedge b = \min(a, b)$; hence

$$f_{\text{FBM}}^* \in \mathcal{H}((2H+1) \wedge 2, L). \quad (23)$$

The expressions of the within- and between-scale wavelet coefficient spectral densities $\mathbf{D}_{j,u}(\cdot; f, \phi, \psi)$ given in Corollary 2 depend both on d and on the function f^* and will therefore be denoted by $\mathbf{D}_{j,u}(\cdot; d, f^*, \phi, \psi)$ in the sequel. We are going to show, however, that these quantities may be approximated by quantities which depend only on the memory parameter d and $f^*(0)$. Let X have a generalized spectral density $f(\lambda) = |1 - e^{i\lambda}|^{-2d} f^*(\lambda)$ and define

$$\sigma_j^2(d, f^*) \stackrel{\text{def}}{=} \text{Var}[W_{j,0}^X] = \int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) |H_j(\lambda)|^2 d\lambda, \quad (24)$$

the variance of the wavelet coefficient of the process X at scale j .

Theorem 3. Let $M \geq 1$ be an integer and α, L, β be constants such that $\alpha > 1$, $0 < L < \infty$ and $\beta \in (0, 2]$. Assume that (W-1)-(W-4) hold with M and α .

(a) Let d_{\min} and d_{\max} be two constants such that

$$[d_{\min}, d_{\max}] \subset ((1 + \beta)/2 - \alpha, M + 1/2). \quad (25)$$

Then, there exists a constant $C > 0$ (only depending on the constants β , d_{\min} , d_{\max} and the functions ϕ and ψ) such that, for all $j \geq 0$, $d \in [d_{\min}, d_{\max}]$ and $f^* \in \mathcal{H}(\beta, L)$,

$$|\sigma_j^2(d, f^*) - f^*(0) K(d, \psi) 2^{2jd}| \leq C f^*(0) L 2^{(2d-\beta)j} \quad (26)$$

where $K(d, \psi)$ is given by

$$K(d, \psi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} |\xi|^{-2d} |\widehat{\psi}(\xi)|^2 d\xi. \quad (27)$$

(b) Let d_{\min} and d_{\max} be two constants such that

$$[d_{\min}, d_{\max}] \subset ((1 + \beta)/2 - \alpha, M]. \quad (28)$$

Then, for all $u \geq 0$, there exists $C > 0$ (only depending on u and on the constants β , d_{\min} , d_{\max} and the functions ϕ and ψ) such that, for all $\lambda \in (-\pi, \pi)$, $j \geq u$, $d \in [d_{\min}, d_{\max}]$ and $f^* \in \mathcal{H}(\beta, L)$,

$$|\mathbf{D}_{j,u}(\lambda; d, f^*, \phi, \psi) - f^*(0) \mathbf{D}_{\infty,u}(\lambda; d, \psi) 2^{2jd}| \leq C f^*(0) L 2^{(2d-\beta)j} \quad (29)$$

where $|\cdot|$ denotes the Euclidean norm in any dimension and, for all $u \geq 0$,

$$\mathbf{D}_{\infty,u}(\lambda; d, \psi) \stackrel{\text{def}}{=} \sum_{l \in \mathbb{Z}} |\lambda + 2l\pi|^{-2d} \mathbf{e}_u(\lambda + 2l\pi) \overline{\widehat{\psi}(\lambda + 2l\pi)} \widehat{\psi}(2^{-u}(\lambda + 2l\pi)). \quad (30)$$

The function $(\lambda, d) \mapsto \mathbf{D}_{\infty,u}(\lambda; d, \psi)$ is (2π) -periodic in λ and jointly continuous on $\mathbb{R} \times [d_{\min}, d_{\max}]$. When $u = 0$, $\mathbf{D}_{\infty,0}$ is a scalar and

$$\int_{-\pi}^{\pi} \mathbf{D}_{\infty,0}(\lambda; d, \psi) d\lambda = K(d, \psi) \neq 0. \quad (31)$$

The proof, based on approximating the wavelet filter transfer function, can be found in Section 6. In order to shed light on Theorem 3, we conclude this section with a number of remarks.

Remark 2. Theorem 3 states that $f^*(0) K(d, \psi) 2^{2jd}$ is a good approximation for $\text{Var}[W_{j,0}^X]$ and that for any $u \geq 0$, $f^*(0) \mathbf{D}_{\infty,u}(\lambda; d, \psi) 2^{2jd}$ is a good $L^\infty(-\pi, \pi)$ approximation to the spectral density $\mathbf{D}_{j,u}(\lambda; d, f^*, \phi, \psi)$.

Remark 3. Relation (29) with $u = 0$ implies (26) since $\left| \int_{-\pi}^{\pi} g_1(\lambda) d\lambda - \int_{-\pi}^{\pi} g_2(\lambda) d\lambda \right| \leq 2\pi \|g_1 - g_2\|_\infty$. Observe, however, that (26) is valid under Condition (25), which is weaker than (28).

Remark 4. Under Condition (28), for all $p > 0$,

$$\|\mathbf{D}_{\infty,d}(\cdot; d, \psi)\|_p \stackrel{\text{def}}{=} \left(\int_{-\pi}^{\pi} |\mathbf{D}_{\infty,d}(\lambda; d, \psi)|^p d\lambda \right)^{1/p} \quad (32)$$

is positive, finite and continuous in $d \in [d_{\min}, d_{\max}]$. This follows from joint continuity and (31).

Remark 5. The spectral density $f^*(0) \mathbf{D}_{\infty,u}(\lambda; d, \psi) 2^{2jd}$, $\lambda \in (-\pi, \pi)$, is in fact the spectral density of the wavelet coefficient of the generalized fractional Brownian motion $B_{(d)}$, where $d \in \mathbb{R}$. The process $B_{(d)}$ is parameterized by a family $\Theta_{(d)}$ of smooth “test” functions $\theta(t)$, $t \in \mathbb{R}$ and is defined as follows: $\{B_{(d)}(\theta), \theta \in \Theta_{(d)}\}$ is a mean zero Gaussian process with covariance

$$\text{Cov} (B_{(d)}(\theta_1), B_{(d)}(\theta_2)) = \int_{\mathbb{R}} |\lambda|^{-2d} \widehat{\theta}_1(\lambda) \overline{\widehat{\theta}_2(\lambda)} d\lambda. \quad (33)$$

The finiteness of the integral $\int_{\mathbb{R}} |\lambda|^{-2d} |\widehat{\theta}(\lambda)|^2 d\lambda$ provides a constraint on the family $\Theta_{(d)}$. For instance, when $d > 1/2$, this condition requires that $\widehat{\theta}(\lambda)$ decays sufficiently quickly at the origin and, when $d < 0$, it requires that $\widehat{\theta}(\lambda)$ decreases sufficiently rapidly at infinity. Hence θ can be a wavelet ψ if $d \in (1/2 - \alpha, M + 1/2)$, which corresponds to (25) with $\beta = 0$. The discrete wavelet transform of $B_{(d)}$ is defined as

$$W_{j,k}^{(d)} \stackrel{\text{def}}{=} B_{(d)}(\psi_{j,k}), \quad j \in \mathbb{Z}, k \in \mathbb{Z}. \quad (34)$$

The spectral density $f^*(0) \mathbf{D}_{\infty,u}(\lambda; d, \psi) 2^{2jd}$ in (29) which serves as an approximation to $\mathbf{D}_{j,u}(\lambda; d, f^*, \phi, \psi)$ is in fact, up to the multiplicative constant $f^*(0)$, the cross spectral density between the wavelet coefficients $W_{j,k}^{(d)}$ and the vector of wavelet coefficients $\mathbf{W}_{j,k}^{(d)}(u) \stackrel{\text{def}}{=} [W_{j-u,2^u k}^{(d)}, \dots, W_{j-u,2^u k+2^u-1}^{(d)}]$. Indeed, using (33), (34) and $\widehat{\psi}_{j,k}(\lambda) = 2^{j/2} \widehat{\psi}(2^j \lambda) e^{-ik\lambda 2^j}$, one has

$$\begin{aligned} \text{Cov} (W_{j,k}^{(d)}, \mathbf{W}_{j,k'}^{(d)}(u)) &= 2^j \int_{\mathbb{R}} |\lambda|^{-2d} \mathbf{e}_u(-2^j \lambda) \widehat{\psi}(2^j \lambda) \overline{\widehat{\psi}(2^{j-u} \lambda)} e^{i\lambda 2^j (k'-k)} d\lambda \\ &= 2^{2dj} \int_{-\pi}^{\pi} \mathbf{D}_{\infty,u}(\lambda; d, \psi) e^{i\lambda(k-k')} d\lambda, \end{aligned} \quad (35)$$

where the last equality is obtained by the change of variable $-2^j \lambda \rightarrow \lambda$ and by folding the integral on $(-\pi, \pi)$.

The within- and between-scale spectral densities $\mathbf{D}_{j,u}(\lambda; d, f^*, \phi, \psi)$ of the process X with memory parameter d may thus be approximated by the DWT of the generalized FBM $B_{(d)}$ (viewed as a generalized process), with an L^∞ error bounded by the RHS in (29).

Remark 6. When d belongs to $(1/2, 3/2)$, $B_{(d)}$ is related to $B_H(t)$, $t \in \mathbb{R}$, by setting $H = d - 1/2 \in (0, 1)$ and, up to a multiplicative constant,

$$B_{(d)}(\theta) = \int_{\mathbb{R}} B_H(t) \theta(t) dt,$$

where the equality holds in the sense of finite-dimensional distributions and hence $\{W_{j,k}^{(d)}, j, k \in \mathbb{Z}\}$ has same distribution as $\{\int_{-\infty}^{\infty} B_H(s) \psi_{j,k}(s) ds, j, k \in \mathbb{Z}\}$. It follows from the previous remark that, for such d and H , the spectral density of the wavelet coefficients of X can be approximated by that of continuous-time fractional Brownian motion.

Remark 7. Once normalized by 2^{2jd} , which is the order of the variance of the wavelet coefficients at scale j (see (26)), the difference of the spectral densities in (29) is bounded by a constant times $2^{-\beta j}$, a factor which tends to zero exponentially fast as $j \rightarrow \infty$. The rate of the decrease is determined by the smoothness exponent β of f^* .

Remark 8. If $d = 0$ and $\{\psi_{j,k}, k \in \mathbb{Z}, j \in \mathbb{Z}\}$ is an orthonormal system, then (33) and (34) imply

$$\text{Cov}(W_{j,k}^{(0)}, W_{j',k'}^{(0)}) = \int_{\mathbb{R}} \widehat{\psi}_{j,k}(\lambda) \overline{\widehat{\psi}_{j',k'}(\lambda)} d\lambda = 2\pi \int_{\mathbb{R}} \psi_{j,k}(t) \psi_{j',k'}(t) dt$$

which vanishes if $j \neq j'$ or $k \neq k'$. Hence, when the memory parameter $d = 0$ and the wavelets are orthonormal, the wavelet coefficients $\{W_{j,k}^X, k \in \mathbb{Z}\}$ are then asymptotically uncorrelated as $j \rightarrow \infty$ and their asymptotic variance is 2π . Using (35), the corresponding cross spectral density is given by

$$\mathbf{D}_{\infty,u}(\lambda; 0, \psi) = 0 \quad \text{if } u > 0 \quad \text{and} \quad \mathbf{D}_{\infty,0}(\lambda; 0, \psi) = 1, \quad \lambda \in (-\pi, \pi). \quad (36)$$

Remark 9. To understand the presence of the asymptotic form $f^*(0)2^{2jd}\mathbf{D}_{\infty,u}(\lambda; d, \psi)$ in (29), start with $\mathbf{D}_{j,u}(\lambda; f, \phi, \psi)$ in (21), use the $2^j(2\pi)$ -periodicity to replace $\sum_{l=0}^{2^j-1}$ by $\sum_{l=-2^{j-1}+1}^{2^j-1}$, and as $j \rightarrow \infty$, approximate f by the spectral density $f^*(0)|\lambda|^{-2d}$ of $\sqrt{f^*(0)}B_{(d)}$, $H_j(\lambda)$ by its asymptotic approximation $2^{j/2}\widehat{\phi}(\lambda)\widehat{\psi}(2^j\lambda)$ in (76) and $H_{j-u}(\lambda)$ by $2^{(j-u)/2}\widehat{\phi}(\lambda)\widehat{\psi}(2^{j-u}\lambda)$, approximate $\widehat{\phi}(2^{-j}(\lambda + 2l\pi))$ by $\widehat{\phi}(0) = 1$ and approximate $\sum_{l=-2^{j-1}+1}^{2^j-1}$ by $\sum_{l=-\infty}^{\infty}$.

Remark 10. Let us examine how Theorem 3 applies when $X(k) = B_H(k)$, $k \in \mathbb{Z}$, that is, X is a discrete-time version of FBM with Hurst index $H \in (0, 1)$. From (4), (5) and (23), we have $d = H + 1/2 \in (1/2, 3/2)$ and $f^* \in \mathcal{H}((2H + 1) \wedge 2, L)$ for some constant L . The condition on M is then $M > H$ in case (a) and $M \geq H + 1/2$ in case (b). The condition on α is $\alpha > \beta/2 - H = (1 - H) \wedge 1/2$ in both cases, which is satisfied because $\alpha > 1$ and $H \in (0, 1)$. Theorem 3 can therefore be applied irrespectively of the value of H when ψ is a Daubechies wavelet with at least $M = 2$ vanishing moments.

Remark 11. In case of FGN, $d = H - 1/2 \in (-1/2, 1/2)$ and hence, compared to the previous case, M decreases by 1 and α increases by 1. Thus the conditions are $M > H - 1$ in case (a) and $M \geq H - 1/2$ in case (b), that is $M = 1$ works in either case. The condition on α becomes $\alpha > (2 - H) \wedge 3/2$, that is $\alpha > 3/2$ will work for all $H \in (0, 1)$. Since the Daubechies wavelet with $M = 2$ has an $\alpha = 1.3390$ (as given by Formula (7.1.23), Page 225 in Daubechies (1992)²), the condition on α is satisfied only for $H > 0.67$. How should one proceed in a situation where $H \in (0, 1)$ is unknown? There are three alternatives: 1) Use $M \geq 3$ since the condition is satisfied for the Daubechies wavelet with $M \geq 3$ (for which $\alpha > 1.63$). 2) Sum the FGN to get a discrete-time version of FBM as above and use $M \geq 2$. 3) Use $M = 2$ and apply Theorem 3 with a smoothness index $\beta' < \beta$ instead of β , worsening the bound in (29).

Remark 12. When $d < 0$, it is not M but α which should influence the choice of the wavelet. The more negative the value of d , the higher the required value of α . Recall that a high value of α corresponds to a fast decrease of $\widehat{\psi}(\xi)$ as $|\xi| \rightarrow \infty$.

5. ANALYSIS OF THE MEMORY PARAMETER ESTIMATOR BASED ON THE REGRESSION OF THE WAVELET VARIANCE

In this section, we consider a *Gaussian* process X with memory parameter d and generalized spectral density $f(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda)$. Then, for any $K > (d - 1/2)$, the distribution of the K -th order increment process $\Delta^K X$ only depends on d and f^* . We apply Theorem 3 to study the wavelet estimator of the memory parameter d , based on the regression of the scale spectrum $\sigma_j^2(d, f^*)$ with respect to the scale index j . This is reasonable because, for large scale j , $\log \sigma_j^2(d, f^*)$ is approximately an affine function of j with slope $(2 \log 2) d$ (see (26) in Theorem 3). Given n observations X_1, \dots, X_n , $\sigma_j^2(d, f^*)$ can be estimated by the empirical variance

$$\widehat{\sigma}_j^2 \stackrel{\text{def}}{=} n_j^{-1} \sum_{k=0}^{n_j-1} (W_{j,k}^X)^2, \quad (37)$$

where for any j , n_j denotes the number of available wavelet coefficients at scale index j , namely, from (12),

$$n_j = [2^{-j}(n - T + 1) - T + 1], \quad (38)$$

²The α in the table on Page 226 is our α minus 1.

where n is the size of the time series and $[x]$ denotes the integer part of x . An estimator of the memory parameter d is then obtained by regressing the logarithm of the empirical variance $\log(\hat{\sigma}_i^2)$ for a finite number of scale indices $j \in \{J_0, \dots, J_0 + \ell\}$ where J_0 is the lower scale and $1 + \ell \geq 2$ is the number of scales in the regression. For a sample size equal to n , this estimator is well defined for J_0 and ℓ such that $\ell \geq 1$ and $J_0 + \ell \leq J(n)$ where

$$J(n) \stackrel{\text{def}}{=} [\log_2(n - T + 1) - \log_2(T)] \quad (39)$$

is the maximal index j such that $n_j \geq 1$. The regression estimator can be expressed formally as

$$\hat{d}_n(J_0, \mathbf{w}) \stackrel{\text{def}}{=} \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \log(\hat{\sigma}_j^2) , \quad (40)$$

where the vector $\mathbf{w} \stackrel{\text{def}}{=} [w_0, \dots, w_\ell]^T$ satisfies

$$\sum_{i=0}^{\ell} w_i = 0 \quad \text{and} \quad 2 \log(2) \sum_{i=0}^{\ell} i w_i = 1 . \quad (41)$$

One may choose, for example, \mathbf{w} corresponding to the weighted least-squares regression vector, defined by

$$\mathbf{w} = D B (B^T D B)^{-1} \mathbf{b} ,$$

where $B \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & \ell \end{bmatrix}^T$ is the so-called design matrix, D is a definite positive matrix and

$$\mathbf{b} \stackrel{\text{def}}{=} [0 \ (2 \log(2))^{-1}]^T . \quad (42)$$

Ordinary least square regression corresponds to the case where D is the identity matrix.

We now compute a bound of the mean square error and an asymptotic equivalent of the variance of $\hat{d}_n(J_0, \mathbf{w})$ in the usual semi-parametric framework adopted by Robinson and his co-authors for studying Fourier estimators. For the wavelet estimator defined above, these quantities depend primarily on n and on the scale index J_0 , while in the Fourier case, the bounds are generally expressed as functions of n and a *bandwidth* parameter m , equal to the number of discrete Fourier frequencies used. To ease comparison, we will express our results with respect to n and m , where m is the number of wavelet coefficients appearing in $\hat{d}_n(J_0, \mathbf{w})$, namely,

$$m \stackrel{\text{def}}{=} \sum_{j=J_0}^{J_0+\ell} n_j .$$

Since $\sum_{j=J_0}^{J_0+\ell} 2^{-j} = 2^{-J_0}(2-2^{-\ell})$, one gets immediately from (38) that $|m-n2^{-J_0}(2-2^{-\ell})| \leq 2(\ell+1)(T-1)$. Thus $m \rightarrow \infty$ is equivalent to having $n2^{-J_0} \rightarrow \infty$, and, when these conditions hold, we have

$$m(n) \sim n2^{-J_0(n)}(2-2^{-\ell}). \quad (43)$$

The next result provides a bound to the bias $\mathbb{E}[\widehat{d}_n(J_0, \mathbf{w})] - d$ and to the variance $\text{Var}[\widehat{d}_n(J_0, \mathbf{w})]$.

Theorem 4. *Assume that (W-1)-(W-4) hold with $M \geq 1$, $\alpha > 1$ and that X is Gaussian. Let \mathbf{w} be a vector satisfying (41) for some $\ell \geq 1$. Let d_{\min} , d_{\max} be two scalars such that $d_{\min} < d_{\max}$ and $[d_{\min}, d_{\max}] \subset ((1+\beta)/2 - \alpha, M]$, where $\beta \in (0, 2]$. Then, there exists a finite constant C (depending only on \mathbf{w} , β , L , d_{\min} , d_{\max} , ϕ and ψ) such that for all $J_0 \in \{0, \dots, J(n) - \ell\}$, $d \in [d_{\min}, d_{\max}]$, and $f^* \in \mathcal{H}(\beta, L)$ with $f^*(0) > 0$*

$$\left| \mathbb{E}[\widehat{d}_n(J_0, \mathbf{w})] - d \right| \leq C \left\{ \left(\frac{m}{n} \right)^\beta + m^{-1} \right\}, \quad (44)$$

$$\text{Var}[\widehat{d}_n(J_0, \mathbf{w})] \leq C \left\{ m^{-1} + \mathbb{1} \left(\frac{m}{n} \geq C^{-1} \right) \right\}. \quad (45)$$

Remark 13. While the bias term bound contains $(m/n)^\beta$, the variance bound has an indicator function which is zero for sufficiently small values of m/n , hence is $o((m/n)^{2\beta})$. This indicator function cannot be dispensed with. Indeed if we start our estimation at the finest scale $J_0 = 0$, corresponding roughly to $m = n$, all we can say is that $\text{Var}(\widehat{d}_n) \leq C$. If, however, we start our estimation at a coarse enough scale J_0 , corresponding to $m/n < C^{-1}$, then $\text{Var}(\widehat{d}_n)$ is bounded by Cm^{-1} , which tends to zero as $m \rightarrow \infty$.

By combining (44) and (45) it is possible to obtain a bound on the mean square error of $\widehat{d}_n(J_0, \mathbf{w})$. More precisely, there exists a constant C (depending only on M , α , β , L , d_{\min} and d_{\max}) such that, for any $f^* \in \mathcal{H}(\beta, L)$, $d \in [d_{\min}, d_{\max}]$ and $J_0 \in \{0, \dots, J(n) - \ell\}$,

$$\mathbb{E} \left[\left\{ \widehat{d}_n(J_0, \mathbf{w}) - d \right\}^2 \right] \leq C \left\{ \left(\frac{m}{n} \right)^{2\beta} + m^{-1} \right\}. \quad (46)$$

This shows in particular that, for any non-decreasing sequence $\{J_0(n), n \geq 0\}$ such that $m^{-1} + m/n \rightarrow 0$, $\widehat{d}_n(\mathbf{w}) \stackrel{\text{def}}{=} \widehat{d}_n(J_0(n), \mathbf{w})$ is a consistent estimator of d . If the regularity exponent β is known, it is possible to choose $J_0(n)$ to balance these two terms, that is, set $(m/n)^{2\beta} \asymp m^{-1}$ or equivalently $2^{J_0(n)} \asymp n^{1/(1+2\beta)}$ as $n \rightarrow \infty$. If we choose $J_0(n)$ in such a

way, (43) and (46) imply

$$\limsup_{n \rightarrow \infty} \sup_{d \in [d_{\min}, d_{\max}]} \sup_{f^* \in \mathcal{H}(\beta, L)} n^{2\beta/(1+2\beta)} \mathbb{E} \left[\left\{ \widehat{d}_n(\mathbf{w}) - d \right\}^2 \right] < \infty.$$

As shown in Giraitis et al. (1997), $n^{-2\beta/(1+2\beta)}$ is the minimax rate of convergence for the memory parameter d in this semi-parametric setting. Therefore,

Corollary 5. *The wavelet estimator is rate optimal in the minimax sense.*

We shall now obtain the asymptotic behavior of $\text{Var} \left[\widehat{d}_n(\mathbf{w}) \right]$ as $n \rightarrow \infty$.

Theorem 6. *Assume that (W-1)-(W-4) hold with $M \geq 1$, $\alpha > 1$ and that X is Gaussian. Let \mathbf{w} be a vector satisfying (41) for some $\ell \geq 1$. Let $\{J_0(n), n \in \mathbb{N}\}$ be a sequence such that $m \rightarrow \infty$ as $n \rightarrow \infty$. For any $f^* \in \mathcal{H}(\beta, L)$, where $\beta \in (0, 2]$, and $d \in ((1 + \beta)/2 - \alpha, M]$,*

$$\lim_{n \rightarrow \infty} m \text{Var} \left[\widehat{d}_n(\mathbf{w}) \right] = (2 - 2^{-\ell}) \mathbf{w}^T \mathbf{V}(d, \psi) \mathbf{w}, \quad (47)$$

where $\mathbf{V}(d, \psi)$ is the $(1 + \ell) \times (1 + \ell)$ matrix defined as

$$\mathbf{V}_{i,j}(d, \psi) \stackrel{\text{def}}{=} \frac{4\pi 2^{2d|i-j|} 2^{i \wedge j}}{K(d, \psi)^2} \int_{-\pi}^{\pi} |\mathbf{D}_{\infty, |i-j|}(\lambda; d, \psi)|^2 d\lambda \quad 0 \leq i, j \leq \ell. \quad (48)$$

Remark 14. The asymptotic expression of the variance (47) is a quadratic form of \mathbf{w} defined by the matrix $\mathbf{V}(d, \psi)$, which depends only on d and ψ (see (48)). The standard theory of linear regression shows that, for any $\ell \geq 1$, the optimal regression vector of length $\ell + 1$ is

$$\mathbf{w}^{\text{opt}}(d, \psi) \stackrel{\text{def}}{=} \mathbf{V}^{-1}(d, \psi) B (B^T \mathbf{V}^{-1}(d, \psi) B)^{-1} \mathbf{b}$$

and the associated limiting variance is $(2 - 2^{-\ell}) \mathbf{b}^T (B^T \mathbf{V}^{-1}(d, \psi) B)^{-1} \mathbf{b}$. This optimal regression vector cannot be used directly since it depends on d which is unknown, but one may apply a two-step procedure using a preliminary estimate of d as in Bardet (2002) in a similar context.

Remark 15. When computing confidence intervals in practice, one sometimes uses asymptotic variances in (47) with $\mathbf{V}^{-1}(0, \psi)$ instead of $\mathbf{V}^{-1}(d, \psi)$, see e.g. Abry and Veitch (1998). The expression $\mathbf{V}^{-1}(0, \psi)$ can be easily obtained if the wavelets are orthonormal. In this case, by (36) and (48), for $i \neq j$, $\mathbf{V}_{i,j}(0) = 0$ and $\mathbf{V}_{j,j}(0) = 8\pi^2 2^j / K(0)^2 = 2^{j+1}$ since by (27), $K(0) = \int_{\mathbb{R}} |\widehat{\psi}(\xi)|^2 d\xi = 2\pi \int_{\mathbb{R}} |\psi(t)|^2 dt = 2\pi$. Then (47) becomes

$$\lim_{n \rightarrow \infty} m \text{Var} \left[\widehat{d}_n(\mathbf{w}) \right] = (2 - 2^{-\ell}) 2 \sum_{j=0}^{\ell} w_j^2 2^j.$$

One can reformulate this in terms of $n_{J_0} \sim n2^{-J_0}$ instead of m . In view of (38) and (43), one gets the following simple expression of the asymptotic variance when $d = 0$:

$$\lim_{n \rightarrow \infty} n_{J_0(n)} \text{Var} \left[\widehat{d}_n(\mathbf{w}) \right] = 2 \sum_{j=0}^{\ell} w_j^2 2^j .$$

If we choose $m(n)$ (or $J_0(n)$) such that the bias in (44) is asymptotically negligible, then we can obtain the asymptotic behavior of the mean square error $\mathbb{E} \left(\widehat{d}_n(\mathbf{w}) - d \right)^2$. In view of (44) and (47), we need $m \rightarrow \infty$ and $\{(m/n)^\beta + m^{-1}\}^2 \ll m^{-1}$, or equivalently

$$n2^{-J_0(n)(1+2\beta)} + n^{-1}2^{J_0(n)} \rightarrow 0, \quad n \rightarrow \infty . \quad (49)$$

Corollary 7. *If (49) holds, then for $f^* \in \mathcal{H}(\beta, L)$ and $d \in ((1 + \beta)/2 - \alpha, M]$,*

$$\lim_{n \rightarrow \infty} n2^{-J_0(n)} \mathbb{E} \left(\widehat{d}_n(\mathbf{w}) - d \right)^2 = \mathbf{w}^T \mathbf{V}(d, \psi) \mathbf{w} .$$

This result of course hints at the existence of a central limit theorem for the estimator $\widehat{d}_n(\mathbf{w})$. Such a result can be obtained by using a central limit theorem for quadratic forms of Gaussian variables which is established in a companion paper Moulines, Roueff and Taqqu (2005).

6. PROOF OF THEOREM 3

From now on, we denote by C constants possibly depending on u , d_{\min} , d_{\max} , β , ϕ and ψ , which may change from line to line and we omit the dependence on ϕ and ψ in the notations. We assume, without loss of generality that $f^*(0) = 1$.

Proof of (a). In the expression (24) of $\sigma_j^2(d, f^*)$, $j \geq 0$, we will approximate $|H_j(\lambda)|^2$ using (79). Thus define

$$A_j \stackrel{\text{def}}{=} 2^j \int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) |\widehat{\phi}(\lambda) \widehat{\psi}(2^j \lambda)|^2 d\lambda \quad \text{and} \quad R_j \stackrel{\text{def}}{=} \sigma_j^2(d, f^*) - A_j .$$

By (79), we have

$$|R_j| \leq C 2^{j(1+M-\alpha)} \int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) |\lambda|^{2M} (1 + 2^j |\lambda|)^{-\alpha-M} d\lambda . \quad (50)$$

We consider A_j and R_j separately starting with A_j .

Express A_j as

$$A_j = 2^j \int_{-\pi}^{\pi} g(\lambda) |\lambda|^{-2d} |\widehat{\psi}(2^j \lambda)|^2 d\lambda , \quad (51)$$

where

$$g(\lambda) \stackrel{\text{def}}{=} f^*(\lambda) |\widehat{\phi}(\lambda)|^2 \left| \lambda / (1 - e^{i\lambda}) \right|^{2d}, \quad \lambda \in (-\pi, \pi). \quad (52)$$

Since $\widehat{\phi}$ is infinitely differentiable by (W-1), $\lambda \mapsto |\widehat{\phi}(\lambda)|^2 \left| \lambda / (1 - e^{i\lambda}) \right|^{2d}$ is infinitely differentiable on $[-\pi, \pi]$. Because $f^* \in \mathcal{H}(\beta, L)$ and $f^*(0) = 1$, there exists a constant C such that for all $\lambda \in (-\pi, \pi)$,

$$|g(\lambda) - g(0)| \leq C L |\lambda|^\beta, \quad (53)$$

where $g(0) = 1$ because $\widehat{\phi}(0) = 1$ by Condition (W-1). Moreover, A_j is finite by (6) since g is bounded and $M > d - 1/2$. We shall now replace the function $g(\lambda)$ by the constant $g(0) = 1$ and extend the interval of integration from $[-\pi, \pi]$ to the whole real line in (51). Eqs. (51) and (53) imply

$$\left| A_j - 2^j \int_{-\pi}^{\pi} g(0) |\lambda|^{-2d} |\widehat{\psi}(2^j \lambda)|^2 d\lambda \right| \leq C L 2^j \int_{-\pi}^{\pi} |\lambda|^{\beta-2d} |\widehat{\psi}(2^j \lambda)|^2 d\lambda.$$

First observe that, after a change of variable,

$$2^j \int_{-\pi}^{\pi} |\lambda|^{\beta-2d} |\widehat{\psi}(2^j \lambda)|^2 d\lambda \leq 2^{j(2d-\beta)} \int_{-\infty}^{\infty} \{ |\lambda|^{\beta-2d_{\min}} \vee |\lambda|^{\beta-2d_{\max}} \} |\widehat{\psi}(\lambda)|^2 d\lambda$$

In the RHS of this inequality, using the behavior of $|\widehat{\psi}(\lambda)|$ at infinity and at the origin implied by (W-2) and (W-3) respectively, and because $d_{\max} < M + 1/2$ and $d_{\min} > (1 + \beta)/2 - \alpha$, the integral is a finite constant. We further observe that, by (W-2), since $d_{\min} > 1/2 - \alpha$, we may write

$$2^j \int_{|\lambda| > \pi} |\lambda|^{-2d} |\widehat{\psi}(2^j \lambda)|^2 d\lambda \leq C 2^{j(1-2\alpha)} \int_{|\lambda| > \pi} |\lambda|^{-2(\alpha+d_{\min})} d\lambda$$

which is integrable. Since, by (25), $1 - 2\alpha < 2d - \beta$, there exists a constant C such that

$$|A_j - K(d) 2^{2jd}| \leq C L 2^{(2d-\beta)j}, \quad (54)$$

where $K(d)$ is given by (27).

We now compute a bound for R_j using (50). Note that there exists a constant C such that, for all $\lambda \in (-\pi, \pi)$,

$$f(\lambda) = f^*(\lambda) \left| \frac{\lambda}{1 - e^{i\lambda}} \right|^{2d} |\lambda|^{-2d} \leq C L |\lambda|^{-2d}. \quad (55)$$

Plugging this into (50) and then separating $\lambda < 1$ and $\lambda \geq 1$, we obtain

$$\begin{aligned} R_j &\leq CL2^{2jd}2^{-j(M+\alpha)} \int_0^{2^j\pi} \{\lambda^{2(M-d_{\min})} \vee \lambda^{2(M-d_{\max})}\} (1+\lambda)^{-\alpha-M} d\lambda \\ &\leq CL2^{j(2d-\beta)}2^{-j(M+\alpha-\beta)} \left\{ \int_0^1 \lambda^{2(M-d_{\max})} d\lambda + \int_1^{2^j\pi} \lambda^{M-2d_{\min}-\alpha} d\lambda \right\}. \end{aligned}$$

Since $2(M - d_{\max}) > -1$, the first integral is a finite constant. Depending on whether $M - 2d_{\min} - \alpha$ is less than, equal to or larger than -1 the second integral is bounded by a finite constant, $\log \pi + j \log 2$ or $C2^{j(1+M-2d_{\min}-\alpha)}$. In the two first cases, we simply observe that $M \geq 1$, $\alpha > 1$ and $\beta \leq 2$ imply $M + \alpha - \beta > 0$, and in the last case that $-(M + \alpha - \beta) + 1 + M - 2d_{\min} - \alpha = 1 - 2d_{\min} - 2\alpha + \beta \leq 0$ by (25) so that, in all cases, $R_j \leq CL2^{(2d-\beta)j}$. This condition, with (54), implies

$$|\sigma_j^2(d, f^*) - K(d)2^{2jd}| = |A_j + R_j - K(d)2^{2jd}| \leq CL2^{(2d-\beta)j}$$

which proves (26).

Proof of (b). For ease of notation, we only consider the case $u = 0$ so that $\mathbf{e}_u(\xi) = 1$. It is also enough to suppose $j \geq 1$. In (21), the summands are $2^j(2\pi)$ -periodic; hence, omitting the summands, $\sum_{l=0}^{2^j-1} = \sum_{l=0}^{2^{j-1}-1} + \sum_{j=2^{j-1}}^{2^j-1} = \sum_{l=0}^{2^{j-1}-1} + \sum_{l=-2^{j-1}}^{-1} = \sum_{l=-2^{j-1}}^{2^{j-1}-1}$. Note that, for $l \in \{-2^{j-1}, \dots, 2^{j-1}-1\}$ and $\lambda \in (0, \pi)$, we have $2^{-j}(\lambda + 2l\pi) \in (-\pi, \pi)$ so that (79) applies. Hence, $\mathbf{D}_{j,0}(\lambda; d, f^*)$ in (21) is expressed as the sum of two functions $A_j(\lambda) + R_j(\lambda)$, defined for all $\lambda \in (0, \pi)$ by

$$A_j(\lambda) \stackrel{\text{def}}{=} \sum_{l=-2^{j-1}}^{2^{j-1}-1} |2^{-j}(\lambda + 2l\pi)|^{-2d} g(2^{-j}(\lambda + 2l\pi)) |\widehat{\psi}(\lambda + 2l\pi)|^2 \quad (56)$$

where g is defined in (52) and where by (79),

$$R_j(\lambda) \leq CL2^{j(2d-M-\alpha)} \sum_{l=-2^{j-1}}^{2^{j-1}-1} |\lambda + 2l\pi|^{2(M-d)} (1 + |\lambda + 2l\pi|)^{-\alpha-M}. \quad (57)$$

From (53), we get, for all $\lambda \in (0, \pi)$,

$$\left| A_j(\lambda) - 2^{2dj} g(0) \sum_{l=-2^{j-1}}^{2^{j-1}-1} |\lambda + 2l\pi|^{-2d} |\widehat{\psi}(\lambda + 2l\pi)|^2 \right| \leq CL2^{(2d-\beta)j} B_j(\lambda), \quad (58)$$

where, by (6) and (W-2), for all $\lambda \in (0, \pi)$,

$$\begin{aligned}
B_j(\lambda) &\stackrel{\text{def}}{=} \sum_{l=-2^{j-1}}^{2^{j-1}-1} |\lambda + 2l\pi|^{\beta-2d} |\widehat{\psi}(\lambda + 2l\pi)|^2 \\
&\leq C \left(|\lambda|^{\beta+2(M-d)} + 2 \sum_{l \geq 1} |\lambda + 2l\pi|^{\beta-2d-2\alpha} \right) \\
&\leq C \left(1 + 2 \sum_{l \geq 1} (2l-1)^{\beta-2d_{\min}-2\alpha} \right) < \infty
\end{aligned} \tag{59}$$

since $|\lambda + 2l\pi| \geq \pi(2l-1)$, $\beta > 0$ and, by (28), one has $M \geq d$ and $\beta - 2d_{\min} - 2\alpha < -1$. By the same arguments, for all $\lambda \in (0, \pi)$,

$$\sum_{|l| \geq 2^{j-1}-1} |\lambda + 2l\pi|^{-2d} |\widehat{\psi}(\lambda + 2l\pi)|^2 \leq C 2^{j(1-2(d_{\min}+\alpha))}$$

is bounded since the exponent is negative. Eqs. (30) with $u = 0$, (58), $g(0) = 1$ and the above inequalities yield that, for all $\lambda \in (0, \pi)$,

$$|A_j(\lambda) - \mathbf{D}_{\infty,0}(\lambda; d) 2^{2dj}| \leq C L 2^{(2d-\beta)j}.$$

We now turn to bounding $R_j(\lambda)$ using (57). For all $\lambda \in (0, \pi)$, using $|\lambda - 2l\pi| \geq \pi(2l-1)$ and (28),

$$R_j(\lambda) \leq C L 2^{j(2d-\beta)} 2^{-j(M+\alpha-\beta)} \left(1 + \sum_{l=1}^{2^j} l^{-2d_{\min}+M-\alpha} \right)$$

which can be bounded as in the proof of (a), by considering the cases $M - 2d_{\min} - \alpha < =$ or > -1 .

The joint continuity of $(\lambda, d) \mapsto \mathbf{D}_{\infty,0}(\lambda; d, \psi)$ on $\mathbb{R} \times [d_{\min}, d_{\max}]$ follows from (59) and dominated convergence.

7. PROOFS OF THEOREM 4 AND THEOREM 6

From now on, we denote by J_{\min} , n_{\min} , C and C' some positive constants whose values may change upon each occurrence and which depend at most on \mathbf{w} , β , L , d_{\min} , d_{\max} , ϕ , and ψ . We will repeatedly use that, by (38), (39) and (43), for $J_0(n) \in \{0, \dots, J(n) - \ell\}$,

$$n_{J_0(n)} \asymp n_{J_0(n)+\ell} \asymp n 2^{-J_0(n)} \asymp m(n), \tag{60}$$

where $a \asymp b$ means that there is a constant $C \geq 1$ such that $a/C \leq b \leq C a$. Finally, for any measurable vector-valued function φ on $[-\pi, +\pi]$ and any $p > 0$, $\|\varphi\|_p = \left(\int_{-\pi}^{\pi} |\varphi(\lambda)|^p d\lambda\right)^{1/p}$.

Proposition 8. *Let $\widehat{\sigma}_j^2$ be defined as in (37) and $\mathbf{D}_{\infty,u}(\lambda; d, \psi)$, $u \geq 0$, be defined as in (30). Under the assumptions of Theorem 4, one has, as $j \rightarrow \infty$ and $n_j \rightarrow \infty$,*

$$2^{-4dj} n_{j-u} \text{Cov}(\widehat{\sigma}_j^2, \widehat{\sigma}_{j-u}^2) \rightarrow (f^*(0))^2 4\pi \|\mathbf{D}_{\infty,u}(\cdot; d, \psi)\|_2^2,$$

uniformly on $d \in [d_{\min}, d_{\max}]$ and $f^* \in \mathcal{H}(\beta, L)$.

Proof. We set $f^*(0) = 1$ without loss of generality. Using (37) and (20), we write

$$\begin{aligned} \text{Cov}[\widehat{\sigma}_j^2, \widehat{\sigma}_{j-u}^2] &= \frac{1}{n_j n_{j-u}} \sum_{k,l=0}^{n_j-1} \sum_{v=0}^{2^u-1} \text{Cov}[W_{j,k}^2, W_{j-u, l2^u+v}^2] \\ &= \frac{2}{n_j n_{j-u}} \sum_{k,l=0}^{n_j-1} |\text{Cov}[W_{j,k}, \mathbf{W}_{j,l}(u)]|^2 \end{aligned} \quad (61)$$

$$= \frac{2}{n_{j-u}} \sum_{\tau \in \mathbb{Z}} \left(1 - \frac{|\tau|}{n_j}\right)_+ |\text{Cov}[W_{j,0}, \mathbf{W}_{j,\tau}(u)]|^2, \quad (62)$$

where, in (61), we used the fact that if the scalar X and the vector $\mathbf{Y} = [Y_1 \dots Y_p]^T$ are jointly Gaussian,

$$\sum_{k=1}^p \text{Cov}(X^2, Y_k^2) = 2 \sum_{k=1}^p \text{Cov}^2(X, Y_k) = 2 |\text{Cov}(X, \mathbf{Y})|^2.$$

Using the notation M_n defined in (81), we have

$$\left(\sum_{\tau \in \mathbb{Z}} \left(1 - \frac{|\tau|}{n_j}\right)_+ |\text{Cov}[W_{j,0}, \mathbf{W}_{j,\tau}(u)]|^2 \right)^{1/2} = M_{n_j}(\mathbf{D}_{j,u}(\cdot; d, f^*)), \quad (63)$$

since, by Corollary 2, $\mathbf{D}_{j,u}(\cdot; d, f^*)$ is the cross-spectral density of the vector $[W_{j,0}, \mathbf{W}_{j,\tau}(u)]$. Applying Lemma 11–(82), the relation $\|\cdot\|_2 \leq \sqrt{2\pi} \|\cdot\|_\infty$ and Theorem 3–(29), there is a constant C such that

$$|M_{n_j}(\mathbf{D}_{j,u}(\cdot; d, f^*)) - 2^{2jd} M_{n_j}(\mathbf{D}_{\infty,u}(\cdot; d))| \leq C 2^{(2d-\beta)j}. \quad (64)$$

On the other hand, by Lemma 11–(83), we have, as $n_j \rightarrow \infty$,

$$|M_{n_j}(\mathbf{D}_{\infty,u}(\cdot; d))|^2 \rightarrow 2\pi \|\mathbf{D}_{\infty,u}(\cdot; d)\|_2^2. \quad (65)$$

The convergence in (65) holds uniformly on $d \in [d_{\min}, d_{\max}]$ because of the joint continuity of $(\lambda, d) \mapsto \mathbf{D}_{\infty, u}(\lambda; d)$ stated in Theorem 3.

The result follows from (62)–(65). \square

Proof of Theorem 4. Again, we set $f^*(0) = 1$ without loss of generality. The bias $\mathbb{E}[\widehat{d}_n(J_0, \mathbf{w})] - d$ can be decomposed into two terms as follows

$$\begin{aligned} \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \mathbb{E} [\log(\widehat{\sigma}_j^2)] - d &= \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \log [\sigma_j^2(d, f^*)] - d + \\ &\quad \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \{ \mathbb{E} [\log(\widehat{\sigma}_j^2)] - \log [\mathbb{E}[\widehat{\sigma}_j^2]] \} , \end{aligned} \quad (66)$$

where $\widehat{\sigma}_j^2$ is the wavelet coefficient empirical variance (37) and $\mathbb{E}[\widehat{\sigma}_j^2] = \sigma_j^2(d, f^*)$.

Using (41), the first term on the RHS of (66) may be rewritten as

$$\sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \log [\sigma_j^2(d, f^*)] - d = \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \log \left(1 + \frac{\sigma_j^2(d, f^*) - K(d) 2^{2jd}}{K(d) 2^{2jd}} \right) \quad (67)$$

By Theorem 3–(26) and using that $\inf_{d \in [d_{\min}, d_{\max}]} K(d) > 0$, there exists a constant C such that

$$\frac{|\sigma_j^2(d, f^*) - K(d) 2^{2jd}|}{K(d) 2^{2jd}} \leq C 2^{-\beta j}.$$

Using that $|\log(1+x)| \leq 2|x|$ for $x \in (-1/2, \infty)$, there is a J_{\min} such that, for $j \geq J_{\min}$, the logarithm in the RHS of (67) is bounded by $C 2^{-\beta j}$, and by (60), for all $J_0 \geq J_{\min}$,

$$\left| \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \log[\sigma_j^2(d, f^*)] - d \right| \leq C \sum_{j=0}^{\ell} |w_j| 2^{-(J_0+j)\beta} \leq C \left(\frac{m}{n} \right)^{\beta}. \quad (68)$$

This bound is in fact valid for all $J_0 \geq 0$ because $\sigma_j^2(d, f^*)$ is bounded away from zero and infinity independently of d and f^* . Indeed, by (24) and since $f^* \in \mathcal{H}(\beta, L)$ with $f^*(0) = 1$, there is a small enough $\epsilon > 0$ only depending on d_{\min} , d_{\max} , β and L such that

$$\begin{aligned} C \int_{-\pi}^{2\epsilon} |H_j(\lambda)|^2 d\lambda &\leq \int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{-2d} (1 - L|\lambda|^{\beta})_+ |H_j(\lambda)|^2 d\lambda \leq \sigma_j^2(d, f^*) \\ &\leq \int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{-2d} (1 + L|\lambda|^{\beta})_+ |H_j(\lambda)|^2 d\lambda \leq C' \int_{-\pi}^{\pi} |\lambda|^{-2d_{\max}} |H_j(\lambda)|^2 d\lambda. \end{aligned} \quad (69)$$

Observe that the lower bound in the previous display does not vanish since, as stated in Remark 1, $H_j(\lambda)$ is a non-zero trigonometric polynomial for all $j \geq 0$ and that the upper

bound is finite since, by (16), $H_j(\lambda) = O(|\lambda|^M)$. Hence there is a positive constant C such that, for all $j = \{0, \dots, J_{\min}\}$, $C^{-1} \leq \sigma_j^2(d, f^*) \leq C$.

We now consider the second term in the RHS of the display (66). The empirical variance (37) is a quadratic form in the wavelet coefficients at $[W_{j,0}, \dots, W_{j,n_j-1}]$. By Corollary 2, these have spectral density $\mathbf{D}_{j,0}(\cdot; d, f^*)$, given in (21). By Lemma 12, the spectral radius of the covariance matrix $\Gamma_j(d, f^*)$ of the random vector $[W_{j,0}, \dots, W_{j,n_j-1}]$ is bounded by the supremum of the spectral density,

$$\rho[\Gamma_j(d, f^*)] \leq 2\pi \|\mathbf{D}_{j,0}(\cdot; d, f^*)\|_\infty. \quad (70)$$

Applying Proposition 13-(84) with $A = n_j^{-1} I_{n_j}$ and $\Gamma = \Gamma_j(d, f^*)$ and using (70), we get

$$|\mathbb{E}[\log(\hat{\sigma}_j^2)] - \log(\mathbb{E}[\hat{\sigma}_j^2])| \leq 4\pi^2 C \left(1 \wedge n_j^{-1} \frac{\|\mathbf{D}_{j,0}(\cdot; d, f^*)\|_\infty^2}{n_j \text{Var}[\hat{\sigma}_j^2]} \right), \quad (71)$$

where C is a universal constant. Now, by Theorem 3-(29) and by joint continuity of $\mathbf{D}_{\infty,0}(\lambda; d)$,

$$2^{-2dj} \|\mathbf{D}_{j,0}(\cdot; d, f^*)\|_\infty \leq C(1 + \|\mathbf{D}_{\infty,0}(\cdot; d)\|_\infty) \leq C.$$

It follows from Proposition 8 that for $j \geq J_{\min}$ and $n_j \geq n_{\min}$,

$$2^{-4dj} n_j \text{Var}(\hat{\sigma}_j^2) \geq 2\pi \inf_{d \in [d_{\min}, d_{\max}]} \|\mathbf{D}_{\infty,0}(\cdot; d)\|_2^2$$

which is positive by Remark 4. The last two displayed equations imply that for $j \geq J_{\min}$ and $n_j \geq n_{\min}$,

$$\frac{\|\mathbf{D}_{j,0}(\cdot; d, f^*)\|_\infty^2}{n_j \text{Var}[\hat{\sigma}_j^2]} \leq C. \quad (72)$$

Inserting (72) into (71) and using (60), we get that for $J_0 \geq J_{\min}$ and $n_{J_0+\ell} \geq n_{\min}$, and $j = J_0, \dots, J_0 + \ell$,

$$|\mathbb{E}[\log(\hat{\sigma}_j^2)] - \log(\mathbb{E}[\hat{\sigma}_j^2])| \leq C n_j^{-1} \leq C n_{J_0+\ell}^{-1} \leq C m^{-1} \leq C(m^{-1} + (m/n)^\beta). \quad (73)$$

This last bound holds in fact without the preceding restrictions on J_0 and $n_{J_0+\ell}$. To see this, use (71) (with the ‘‘bound 1’’) and observe that, by (60), $J_0 \leq J_{\min}$ implies $2^{-J_0} \geq 2^{-J_{\min}}$, that is, $m/n \geq C$, and $n_{J_0+\ell} \leq n_{\min}$ implies $m^{-1} \geq C$. The bounds (68) and (73), inserted in (66), yield the bound (44) on the bias.

We now compute the variance of the estimator $\hat{d}_n(J_0, \mathbf{w})$. By Proposition 13–(86) and using (70) and (72) as in (71),

$$\begin{aligned}
& \left| \text{Var}(\hat{d}_n(J_0, \mathbf{w})) - \sum_{i,j=J_0}^{J_0+\ell} w_{i-J_0} w_{j-J_0} \frac{\text{Cov}(\hat{\sigma}_i^2, \hat{\sigma}_j^2)}{\mathbb{E}[\hat{\sigma}_i^2] \mathbb{E}[\hat{\sigma}_j^2]} \right| \\
& \leq \sum_{i,j=J_0}^{J_0+\ell} |w_{i-J_0} w_{j-J_0}| \left| \text{Cov}(\log(\hat{\sigma}_i^2), \log(\hat{\sigma}_j^2)) - \frac{\text{Cov}(\hat{\sigma}_i^2, \hat{\sigma}_j^2)}{\mathbb{E}[\hat{\sigma}_i^2] \mathbb{E}[\hat{\sigma}_j^2]} \right| \\
& \leq C \sum_{i,j=J_0}^{J_0+\ell} |w_{i-J_0} w_{j-J_0}| \left\{ \frac{n_i^{-3} \|\mathbf{D}_{i,0}(\cdot; d, f^*)\|_\infty^3}{\text{Var}^{3/2}(\hat{\sigma}_i^2)} \vee \frac{n_j^{-3} \|\mathbf{D}_{j,0}(\cdot; d, f^*)\|_\infty^3}{\text{Var}^{3/2}(\hat{\sigma}_j^2)} \right\} \\
& \leq C n_{J_0+\ell}^{-3/2} \leq C m^{-3/2} = o(m^{-1}) . \tag{74}
\end{aligned}$$

On the other hand, by Proposition 8 and Theorem 3–(26), we have, for any $u \geq 0$, as $j_0 \rightarrow \infty$ and $n_{j_0} \rightarrow \infty$,

$$\frac{n_{j_0-u} \text{Cov}(\hat{\sigma}_{j_0}^2, \hat{\sigma}_{j_0-u}^2)}{\mathbb{E}[\hat{\sigma}_{j_0}^2] \mathbb{E}[\hat{\sigma}_{j_0-u}^2]} = \frac{n_{j_0-u} \text{Cov}(\hat{\sigma}_{j_0}^2, \hat{\sigma}_{j_0-u}^2)}{\sigma_{j_0}^2(d, f^*) \sigma_{j_0-u}^2(d, f^*)} \rightarrow \frac{4\pi \|\mathbf{D}_{\infty,u}(\cdot; d)\|_2^2}{2^{-2du} (\mathbf{K}(d))^2} \tag{75}$$

uniformly in $d \in [d_{\min}, d_{\max}]$ and $f^* \in \mathcal{H}(\beta, L)$. Applying (75) with $j_0 = i \vee j$ and $u = |i-j|$ so that $j_0 - u = i \wedge j$, and since $\inf_{d \in [d_{\min}, d_{\max}]} \mathbf{K}(d) > 0$, for all $J_0 \geq J_{\min}$ and $n_{J_0+\ell} \geq n_{\min}$,

$$\left| \sum_{i,j=J_0}^{J_0+\ell} w_{i-J_0} w_{j-J_0} \frac{\text{Cov}(\hat{\sigma}_i^2, \hat{\sigma}_j^2)}{\sigma_i^2(d, f^*) \sigma_j^2(d, f^*)} \right| \leq C n_{J_0+\ell}^{-1} \leq C m^{-1} .$$

This bound with (74) yields (45) for $J_0 \geq J_{\min}$ and $n_{J_0+\ell} \geq n_{\min}$. When $J_0 \leq J_{\min}$ or $n_{J_0+\ell} \leq n_{\min}$, the RHS of (45) is larger than a positive constant and it suffices to use that, by the Minkowski inequality,

$$\text{Var}^{1/2}(\hat{d}_n(J_0, \mathbf{w})) \leq \sum_{j=J_0}^{J_0+\ell} |w_{j-J_0}| \text{Var}^{1/2}(\log(\hat{\sigma}_j^2)) \leq C ,$$

where we applied Proposition 13–(85) and that, by (69), $\mathbb{E}[\hat{\sigma}_j^2] = \sigma_j^2(d, f^*)$ does not vanish. \square

Proof of Theorem 6. By (38) and (43), for $i, j = J_0, \dots, J_0 + \ell$,

$$n_{i \wedge j}^{-1} \sim (n 2^{-(i \wedge j)})^{-1} \sim m^{-1} (2 - 2^{-l}) 2^{(i - J_0) \wedge (j - J_0)} .$$

Hence, by (74) and (75),

$$\lim_{m \rightarrow \infty} m \text{Var}(\widehat{d}_n(J_0, \mathbf{w})) = (2 - 2^\ell) \sum_{i,j=J_0}^{J_0+\ell} w_{i-J_0} w_{j-J_0} \frac{4\pi 2^{2d|i-j|}}{(\text{K}(d))^2} 2^{(i-J_0) \wedge (j-J_0)} \|\mathbf{D}_{\infty,|i-j|}(\cdot; d)\|_2^2,$$

which gives (47) after a change of variables. \square

APPENDIX A. APPROXIMATION OF WAVELET FILTER TRANSFER FUNCTIONS

Proposition 9. *Under (W-1)-(W-4), there exist positive constants C_i , $i = 1, \dots, 4$ only depending on ϕ and ψ , such that, for all $j \geq 0$ and $\lambda \in (-\pi, \pi)$,*

$$|H_j(\lambda) - 2^{j/2} \widehat{\phi}(\lambda) \overline{\widehat{\psi}(2^j \lambda)}| \leq C_1 2^{j(1/2-\alpha)} |\lambda|^M, \quad (76)$$

$$|\widehat{\phi}(\lambda) \widehat{\psi}(2^j \lambda)| \leq C_2 |2^j \lambda|^M (1 + 2^j |\lambda|)^{-\alpha-M}, \quad (77)$$

$$|H_j(\lambda)| \leq C_3 2^{j/2} |2^j \lambda|^M (1 + 2^j |\lambda|)^{-\alpha-M}, \quad (78)$$

$$|H_j(\lambda)|^2 - 2^j |\widehat{\phi}(\lambda) \widehat{\psi}(2^j \lambda)|^2 \leq C_4 2^{j(1+M-\alpha)} |\lambda|^{2M} (1 + 2^j |\lambda|)^{-\alpha-M}. \quad (79)$$

Proof. Under (W-1) and (W-2), we have that, for all $t \in \mathbb{R}$, $\sum_{k \in \mathbb{Z}} \widehat{\phi}(\lambda + 2k\pi) e^{it(\lambda+2k\pi)}$ is a 2π -periodic function, integrable on $(-\pi, \pi)$ and whose l -th Fourier coefficients is

$$\int_{-\pi}^{\pi} \sum_{k \in \mathbb{Z}} \widehat{\phi}(\lambda + 2k\pi) e^{it(\lambda+2k\pi)} e^{-i\lambda l} d\lambda = \int_{-\infty}^{\infty} \widehat{\phi}(\lambda) e^{it\lambda} e^{-i\lambda l} d\lambda = 2\pi \phi(t - l).$$

It follows that, for all λ and t in \mathbb{R} ,

$$\sum_{l \in \mathbb{Z}} \phi(t - l) e^{i\lambda l} = \sum_{k \in \mathbb{Z}} \widehat{\phi}(\lambda + 2k\pi) e^{it(\lambda+2k\pi)},$$

which is a form of the Poisson summation formula. Inserting this in (14) gives

$$\begin{aligned} H_j(\lambda) &= 2^{-j/2} \int_{-\infty}^{\infty} \left(\sum_{k \in \mathbb{Z}} \widehat{\phi}(\lambda + 2k\pi) e^{it(\lambda+2k\pi)} \right) \psi(2^{-j} t) dt \\ &= 2^{-j/2} \sum_{k \in \mathbb{Z}} \widehat{\phi}(\lambda + 2k\pi) \int_{-\infty}^{\infty} e^{it(\lambda+2k\pi)} \psi(2^{-j} t) dt \\ &= 2^{j/2} \sum_{k \in \mathbb{Z}} \widehat{\phi}(\lambda + 2k\pi) \overline{\widehat{\psi}(2^j(\lambda + 2k\pi))}. \end{aligned}$$

From this expression of H_j , we get, for all $j \geq 0$ and $\lambda \in (-\pi, \pi)$,

$$|H_j(\lambda) - 2^{j/2} \widehat{\phi}(\lambda) \overline{\widehat{\psi}(2^j \lambda)}| = 2^{j/2} \left| \sum_{|k| \geq 1} \widehat{\phi}(\lambda + 2k\pi) \overline{\widehat{\psi}(2^j(\lambda + 2k\pi))} \right|. \quad (80)$$

Now using successively (7) and (W-2), there is a constant C such that, for all non-zero integer k and all $\lambda \in (-\pi, \pi)$, $|\widehat{\phi}(\lambda + 2k\pi)| \leq C|\lambda|^M$ and

$$|\overline{\widehat{\psi}(2^j(\lambda + 2k\pi))}| \leq C(2^j|\lambda + 2k\pi|)^{-\alpha} \leq C2^{-\alpha j}(2|k|\pi - |\lambda|)^{-\alpha} \leq \frac{C2^{-\alpha j}}{\pi^\alpha(2|k| - 1)^\alpha}.$$

Inserting these bounds into (80) gives (76).

The bound (77) follows from (W-1) ($|\widehat{\phi}(\xi)| \leq \int_{-\infty}^{\infty} |\phi(t)|dt < \infty$), (W-2) ($|\widehat{\psi}(\xi)| \leq C(1 + |\xi|)^{-\alpha}$) and (6).

The two last bounds (78) and (79) follow from the two first (76) and (77). Indeed, let $H_j^{(0)}(\lambda) \stackrel{\text{def}}{=} 2^{j/2} \widehat{\phi}(\lambda) \widehat{\psi}(2^j \lambda)$. For (78) we write

$$|H_j(\lambda)| \leq |H_j(\lambda) - H_j^{(0)}(\lambda)| + |H_j^{(0)}(\lambda)|.$$

Applying (76) and (77), the RHS of this equation is bounded by

$$\begin{aligned} C_1 2^{j(1/2-\alpha)} |\lambda|^M + C_2 2^{j/2} |2^j \lambda|^M (1 + 2^j |\lambda|)^{-\alpha-M} \leq \\ 2^{j/2} |2^j \lambda|^M (1 + 2^j |\lambda|)^{-\alpha-M} (C_1 2^{-j(\alpha+M)} (1 + 2^j |\lambda|)^{\alpha+M} + C_2). \end{aligned}$$

By observing that, for all $j \geq 0$ and $\lambda \in (-\pi, \pi)$, the last term in parentheses is bounded by $C_1 2^{-j(\alpha+M)} (2^{1+j} \pi)^{\alpha+M} + C_2 \leq C_1 (2\pi)^{\alpha+M} + C_2$, we get (78). For (79), we write

$$\left| |H_j(\lambda)|^2 - |H_j^{(0)}(\lambda)|^2 \right| \leq \left| H_j(\lambda) - H_j^{(0)}(\lambda) \right| \left(|H_j^{(0)}(\lambda)| + |H_j(\lambda)| \right)$$

and apply (76), (77) and (78). \square

Corollary 10. *Under (W-1)-(W-4), there exists $j_0 \geq 0$ such that, for all $j \geq j_0$, H_j is not identically zero.*

Proof. By (W-1), there exist sufficiently small positive constants ϵ and η such that $|\widehat{\phi}(\lambda)| \geq 1/2$ for all $|\lambda| \leq \epsilon$ and $\inf_{|\lambda| \leq \epsilon^{-1}} |\widehat{\psi}(\lambda)| \geq \eta$. Hence for all j such that $2^j \epsilon \geq \epsilon^{-1}$, using (76),

$$\inf_{|\lambda| \leq \epsilon} 2^{-j/2} |H_j(\lambda)| \geq \inf_{|\lambda| \leq \epsilon} |\widehat{\phi}(\lambda) \widehat{\psi}(2^j \lambda)| - C_1 2^{-j\alpha} \sup_{|\lambda| \leq \epsilon} |\lambda|^M \geq \eta/2 - C_1 2^{-j\alpha} \epsilon^M,$$

which is positive for j large enough. \square

APPENDIX B. SOME USEFUL INEQUALITIES

Lemma 11. *Let p be a positive integer. For all \mathbb{C}^p -valued function $\mathbf{g} \in L^2(-\pi, \pi)$ and $n \geq 1$, define*

$$M_n(\mathbf{g}) \stackrel{\text{def}}{=} \left\{ \sum_{k \in \mathbb{Z}} \left(1 - \frac{|k|}{n} \right)_+ |\mathbf{c}_k|^2 \right\}^{1/2}, \quad (81)$$

where $|\cdot|$ denotes the Euclidean norm in any dimension and $\mathbf{c}_k = \int_{-\pi}^{\pi} \mathbf{g}(\lambda) e^{ik\lambda} d\lambda$. Then, for all \mathbf{g}_1 and \mathbf{g}_2 in $L^2(-\pi, \pi)$,

$$|M_n(\mathbf{g}_1) - M_n(\mathbf{g}_2)| \leq \sqrt{2\pi} \left(\int_{-\pi}^{\pi} |\mathbf{g}_1(\lambda) - \mathbf{g}_2(\lambda)|^2 d\lambda \right)^{1/2}. \quad (82)$$

Moreover, for all \mathbf{g} in $L^2(-\pi, \pi)$, as $n \rightarrow \infty$,

$$M_n(\mathbf{g}) \rightarrow \sqrt{2\pi} \left(\int_{-\pi}^{\pi} |\mathbf{g}(\lambda)|^2 d\lambda \right)^{1/2}. \quad (83)$$

Proof. Suppose $p = 1$ (the proof for $p > 1$ is identical). Observe that the RHS in (81) is a norm on $(c_k)_{k \in \mathbb{Z}} \in l^2(\mathbb{Z})$ which is bounded by the l^2 norm $(\sum_{k \in \mathbb{Z}} |c_k|^2)^{1/2}$. Thus, by Parseval Theorem, $M_n(g)$ is a norm on $g \in L^2(-\pi, \pi)$ which is bounded by $\sqrt{2\pi} \|g\|_2$. Hence $|M_n(g_1) - M_n(g_2)| \leq M_n(g_1 - g_2) \leq \sqrt{2\pi} \|g_1 - g_2\|_2$. Finally, (83) is obtained by dominated convergence. \square

Denote by $\text{Tr}(A)$ and $\rho(A)$ the trace and the spectral radius of a matrix A . Recall that $\rho(A)$ is the maximum of the modulus of the eigenvalues of A .

Lemma 12. *Let $\{\xi_\ell, \ell \in \mathbb{Z}\}$ be a stationary process with spectral density g and let Γ_n be the covariance matrix of $[\xi_1, \dots, \xi_n]$. Then, $\rho(\Gamma_n) \leq 2\pi \|g\|_\infty$.*

Proof. Since Γ_n is a non-negative definite matrix, $\rho(\Gamma_n) = \sup_{\mathbf{x} \in \mathbb{R}^n, |\mathbf{x}| \leq 1} \mathbf{x}^T \Gamma_n \mathbf{x}$, where $|\mathbf{x}|$ is the Euclidean norm of \mathbf{x} . For all $\mathbf{x} \in \mathbb{R}^n$, we may write

$$\mathbf{x}^T \Gamma_n \mathbf{x} = \int_{-\pi}^{\pi} g(\lambda) \left| \sum_{\ell=1}^n \mathbf{x}_\ell e^{-i\ell\lambda} \right|^2 d\lambda \leq \|g\|_\infty \int_{-\pi}^{\pi} \left| \sum_{\ell=1}^n \mathbf{x}_\ell e^{-i\ell\lambda} \right|^2 d\lambda = 2\pi \|g\|_\infty |\mathbf{x}|^2.$$

\square

Proposition 13. *Let ξ is a zero-mean $n \times 1$ Gaussian vector with covariance Γ . Then there exists a universal constant C independent of n such that for any $n \times n$ non-negative*

symmetric matrices A satisfying $\text{Tr}(A\Gamma) > 0$,

$$|\mathbb{E}(\log[\xi^T A \xi]) - \log(\mathbb{E}[\xi^T A \xi])| \leq C \left(1 \wedge \frac{\rho^2(A)\rho^2(\Gamma)}{\text{Var}(\xi^T A \xi)} \right); \quad (84)$$

$$\text{Var}(\log[\xi^T A \xi]) \leq C. \quad (85)$$

Let $[\xi^T, \tilde{\xi}^T]^T$ be a zero-mean $(n + \tilde{n}) \times 1$ Gaussian vector such that $\text{Cov}(\xi) = \Gamma$ and $\text{Cov}(\tilde{\xi}) = \tilde{\Gamma}$. Then there exists a universal constant C independent of n and \tilde{n} such that for any $n \times n$ and $\tilde{n} \times \tilde{n}$ non-negative symmetric matrices A and \tilde{A} satisfying $\text{Tr}(A\Gamma) > 0$ and $\text{Tr}(\tilde{A}\tilde{\Gamma}) > 0$,

$$\left| \text{Cov}(\log[\xi^T A \xi], \log[\tilde{\xi}^T \tilde{A} \tilde{\xi}]) - \frac{\text{Cov}(\xi^T A \xi, \tilde{\xi}^T \tilde{A} \tilde{\xi})}{\mathbb{E}[\xi^T A \xi] \mathbb{E}[\tilde{\xi}^T \tilde{A} \tilde{\xi}]} \right| \leq C \left\{ \frac{\rho^3(A)\rho^3(\Gamma)}{\text{Var}^{3/2}(\xi^T A \xi)} \vee \frac{\rho^3(\tilde{A})\rho^3(\tilde{\Gamma})}{\text{Var}^{3/2}(\tilde{\xi}^T \tilde{A} \tilde{\xi})} \right\}. \quad (86)$$

Proof. Let k be the rank of Γ and Q be $n \times k$ full rank matrix such that $QQ^T = \Gamma$. Let $\zeta \sim \mathcal{N}(0, I_k)$, where I_k is the identity matrix of size $k \times k$. For any unitary matrix U , $U\zeta \sim \mathcal{N}(0, I_k)$ and hence $QU\zeta$ has same distribution as ξ . Moreover, since A is symmetric, so is $Q^T A Q$. We may choose an unitary matrix U such that $\Lambda \stackrel{\text{def}}{=} U^T(Q^T A Q)U$ is a diagonal matrix with non-negative entries. Furthermore,

$$\zeta^T \Lambda \zeta = (QU\zeta)^T A (QU\zeta) \stackrel{\text{d}}{=} \xi^T A \xi, \quad (87)$$

where $\stackrel{\text{d}}{=}$ denotes the equality of distributions. Since Λ is diagonal with non-negative diagonal entries $(\lambda_i)_{i=1,\dots,k}$, $\zeta^T \Lambda \zeta$ is a sum of independent r.v.'s of the form $\sum_{i=1}^k \lambda_i \zeta_i^2$. Since $\mathbb{E}\zeta_i^2 = 1$ and $\text{Var}(\zeta_i^2) = 2$, we get from (87) that $\sum_{i=1}^k \lambda_i = \mathbb{E}[\zeta^T \Lambda \zeta] = \mathbb{E}[\xi^T A \xi] = \text{Tr}(A\Gamma) > 0$ and $\text{Var}[\xi^T A \xi] = \text{Var}[\zeta^T \Lambda \zeta] = 2 \sum_{i=1}^k \lambda_i^2$. Now set

$$S \stackrel{\text{def}}{=} \frac{\xi^T A \xi}{\mathbb{E}[\xi^T A \xi]} \stackrel{\text{d}}{=} \sum_{i=1}^k d_i \zeta_i^2 \quad \text{with} \quad d_i \stackrel{\text{def}}{=} \frac{\lambda_i}{\sum_{j=1}^k \lambda_j}, \quad (88)$$

so that

$$\mathbb{E}[S] = 1 \quad \text{and} \quad \text{Var}(S) = 2 \|d\|^2, \quad (89)$$

where $\|d\|^2 \stackrel{\text{def}}{=} \sum_{i=1}^k d_i^2$. The quantities of interest in (84) and (85) become

$$\mathbb{E}(\log[\xi^T A \xi]) - \log(\mathbb{E}[\xi^T A \xi]) = \mathbb{E}(\log[S]) \quad (90)$$

and

$$\text{Var}(\log[\xi^T A \xi]) = \text{Var}(\log[S]) . \quad (91)$$

Since $\sum_{i=1}^k \lambda_i \geq \sum_{i=1}^k \lambda_i^2 / \max_{1 \leq i \leq k} \lambda_i$ and $\rho(\Lambda) = \max_{1 \leq i \leq k} \lambda_i$, we get

$$\|d\|^2 = \frac{\sum_{i=1}^k \lambda_i^2}{\left(\sum_{i=1}^k \lambda_i\right)^2} \leq \frac{\rho^2(\Lambda)}{\sum_{i=1}^k \lambda_i^2} = \frac{2\rho^2(\Lambda)}{\text{Var}(\xi^T A \xi)} \leq \frac{2\rho^2(A) \rho^2(\Gamma)}{\text{Var}(\xi^T A \xi)} . \quad (92)$$

This is the quantity which appears in (84), and we will therefore express bounds in terms of $\|d\|$.

Denote by F the distribution function of S , that is $F(x) = \mathbb{P}(S \leq x)$. Observe that $F(0) = 0$ since S is a non-negative weighted sum of independent central chi-squares and that all the weights do not vanish. To obtain exponential bounds on F , observe that, by standard computations on the chi-square distribution, one has, for $t > -(2 \max_{1 \leq i \leq k} d_i)^{-1}$,

$$\mathbb{E}[e^{-tS}] = \prod_{i=1}^k \mathbb{E}[e^{-td_i \zeta_i^2}] = \prod_{i=1}^k (1 + 2d_i t)^{-1/2} . \quad (93)$$

Therefore, for any $t > 0$ and $x > 0$,

$$\log[F(x)] \leq \log[\mathbb{E}[e^{-tS}]] = xt - (1/2) \sum_{i=1}^k \log(1 + 2d_i t) . \quad (94)$$

Using (94), we derive two bounds for $F(x)$ by choosing t adequately. One bound, which will not depend on $\|d\|$ is for x around 0, the other one, which will improve as $\|d\|$ decreases, is for x in $(0, 1/2)$.

To get the first bound, observe that, for $t \geq 0$, $\prod_{i=1}^k (1 + 2d_i t) \geq 1 + 2t \sum_{i=1}^k d_i = 1 + 2t$, $\sum_{i=1}^k \log(1 + 2d_i t) \geq \log(1 + 2t)$. Plugging this inequality in (94) and setting $t = 1/(2x)$ yields

$$F(x) \leq e^{1/2} \left(\frac{x}{1+x} \right)^{1/2} \leq \sqrt{e x}, \quad x > 0 . \quad (95)$$

Let $p \geq 1$ and $\alpha \in \mathbb{R}$. Since $\lim_{x \rightarrow 0^+} |\log(x)|^p F(x) = 0$, integration by parts and (95) give that,

$$\int_0^1 |\log(x)|^p dF(x) = p \int_0^1 |\log(x)|^{p-1} x^{-1} F(x) dx \leq e^{1/2} \int_0^1 |\log(x)|^{p-1} x^{-1/2} dx ,$$

which is a finite constant. Since $\sup_{x \geq 1} (|\log(x)|^p/x)$ is finite and $\mathbb{E}S = 1$, we get that $\int_1^\infty |\log(x)|^p dF(x)$ is bounded by a constant only depending on p and thus

$$\mathbb{E}[|\log S|^p] = \int_0^1 |\log(x)|^p dF(x) + \int_1^\infty |\log(x)|^p dF(x) \leq K_p , \quad (96)$$

where K_p is constant only depending on p . This bound proves the left part of the \wedge sign in (84).

We now derive a second bound on $F(x)$ which will yield the right part of the \wedge sign in (84). Since the second derivative of $\log(1+u)$ has absolute value at most 1 for all $u \geq 0$, we have, by Taylor's formula, that, for any $t \geq 0$, $\log(1+2d_i t) \geq 2d_i t - 2d_i^2 t^2$. Applying this to (94) and using $\sum_{i=1}^k d_i = 1$, we get

$$\log[F(x)] \leq (x-1)t + t^2\|d\|^2, \quad x > 0 .$$

Setting this time $t = \|d\|^{-1}$, we obtain the following exponential bound:

$$F(x) \leq \exp[-(1-x)\|d\|^{-1} + 1] \leq \exp[-\|d\|^{-1}/2 + 1] , \quad x \in (0, 1/2) . \quad (97)$$

Using the relation $a \wedge b \leq \sqrt{ab}$, $a, b \geq 0$, we can combine (95) and (97) to get

$$F(x) \leq x^{1/4} \exp(-\|d\|^{-1}/4 + 3/4) , \quad x \in (0, 1/2) .$$

With this last bound of F at hand, we can improve the bound established in (96) as follows. Let $p \geq 1$. Since $|\log(x)|^p x^{1/4}$ is bounded on $x \in (0, 1/2)$ and $|\log(x)|^{p-1} x^{-3/4}$ is integrable on $x \in (0, 1/2)$, we have, by integration by parts,

$$\begin{aligned} \mathbb{E}(|\log(S)|^p \mathbb{1}\{S \leq 1/2\}) &\leq [|\log(x)|^p F(x)]_0^{1/2} + p \int_0^{1/2} |\log(x)|^{p-1} x^{-1} F(x) dx \\ &\leq C_p \exp(-\|d\|^{-1}/4) \leq C_{p,\alpha} \|d\|^\alpha , \end{aligned} \quad (98)$$

where C_p and $C_{p,\alpha}$ are constants only depending on p and (p, α) . For $x \in (0, 1)$, we have $|\log(x) - (x-1)| \leq |\log(x)|$, and for $x \geq 1/2$, a Taylor expansion gives $|\log(x) - (x-1)| \leq 2(x-1)^2$, since the second derivative of $\log(x)$ has absolute value at most $(1/2)^{-2} = 4$. Hence, for any $x > 0$,

$$|\log(x) - (x-1)| \leq |\log(x)| \mathbb{1}_{[0,1/2]}(x) + 2(x-1)^2 \mathbb{1}_{[1/2,\infty)}(x) . \quad (99)$$

Since $\mathbb{E}[S - 1] = 0$ and $\mathbb{E}[(S - 1)^2] = \text{Var}(S) = \|d\|^2 \text{Var}(\zeta_1^2) = 2\|d\|^2$, using (99) and (98) with $p = 1$ and $\alpha = 2$, we get

$$\begin{aligned} |\mathbb{E}[\log(S)]| &= |\mathbb{E}[\log(S) - (S - 1)]| \leq \mathbb{E}[|\log(S)| \mathbb{1}\{|S| \leq 1/2\}] + 2\mathbb{E}[(S - 1)^2] \\ &\leq C_{1,2}\|d\|^2 + 4\|d\|^2. \end{aligned}$$

Applying (90) and (92), we get the inequality (84) with the right part of the \wedge sign.

The bound (85) is obtained by applying (96) since, by (91), $\text{Var}(\log[\xi^T A \xi]) = \text{Var}(\log(S)) \leq \mathbb{E} \log^2 S$.

We now prove (86). Define \tilde{k} , \tilde{d}_i and \tilde{S} as we did k , d_i and S . The LHS in (86) then reads

$$\begin{aligned} \mathbb{E}[\log(S) \log(\tilde{S})] - \mathbb{E}[(S - 1)(\tilde{S} - 1)] &= \mathbb{E}[(S - 1)(\log(\tilde{S}) - (\tilde{S} - 1))] \\ &\quad + \mathbb{E}[(\tilde{S} - 1)(\log(S) - (S - 1))] + \mathbb{E}[(\log(S) - (S - 1))(\log(\tilde{S}) - (\tilde{S} - 1))]. \end{aligned}$$

We will provide a bound for the first term of the RHS of this display, the other terms being treated similarly. By using (99) and the Cauchy-Schwarz inequality,

$$\begin{aligned} &|\mathbb{E}[(S - 1)(\log(\tilde{S}) - (\tilde{S} - 1))]| \\ &\leq \mathbb{E}[(S - 1)|\log(\tilde{S})| \mathbb{1}_{[0,1/2]}(\tilde{S})] + 2\mathbb{E}[(S - 1)(\tilde{S} - 1)^2] \\ &\leq \left(\mathbb{E}|S - 1|^2 \mathbb{E}[|\log(\tilde{S})|^2 \mathbb{1}_{[0,1/2]}(\tilde{S})]\right)^{1/2} + 2\left(\mathbb{E}|S - 1|^2 \mathbb{E}|\tilde{S} - 1|^4\right)^{1/2}. \end{aligned}$$

In view of (92), it remains to show that the two last terms are $O(\|d\|^3 \vee \|\tilde{d}\|^3)$. By definition, $\tilde{S} - 1 = \sum_{i=1}^{\tilde{k}} \tilde{d}_i(\zeta_i^2 - 1)$, where $\{\zeta_i\}_{1 \leq i \leq \tilde{k}}$ are i.i.d. standard normal. Therefore,

$$\mathbb{E}|\tilde{S} - 1|^4 = \sum_{i=1}^{\tilde{k}} \tilde{d}_i^4 \text{cum}_4(\zeta_1^2) + 3 \left(\sum_{i=1}^{\tilde{k}} \tilde{d}_i^2\right)^2 \text{Var}^2(\zeta_1^2),$$

where $\text{cum}_4(Z)$ is the fourth-order cumulant of the random variable Z . Since $\sum_{i=1}^{\tilde{k}} \tilde{d}_i^4 \leq \left(\sum_{i=1}^{\tilde{k}} \tilde{d}_i^2\right)^2$, we obtain that $\mathbb{E}|\tilde{S} - 1|^4 \leq C\|\tilde{d}\|^4$ for some constant C . Therefore,

$$(\mathbb{E}|S - 1|^2)^{1/2} \left(\mathbb{E}|\tilde{S} - 1|^4\right)^{1/2} \leq C\|d\|\|\tilde{d}\|^2 \leq C(\|d\|^3 \vee \|\tilde{d}\|^3).$$

Applying (98) with $p = 2$ and $\alpha = 4$, we have

$$(\mathbb{E}|S - 1|^2)^{1/2} \left(\mathbb{E}[|\log(\tilde{S})|^2 \mathbb{1}_{[0,1/2]}(\tilde{S})]\right)^{1/2} \leq (\sqrt{2}\|d\|)(C_{2,4}\|\tilde{d}\|^2) \leq \sqrt{2}C_{2,4}(\|d\|^3 \vee \|\tilde{d}\|^3),$$

for some constant C' , which concludes the proof. \square

REFERENCES

ABRY, P. and VEITCH, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE Trans. Inform. Theory* **44** 2–15.

BARDET, J.-M. (2002). Statistical study of the wavelet analysis of fractional Brownian motion. *IEEE Trans. Inform. Theory* **48** 991–999.

BARDET, J. M., LANG, G., MOULINES, E. and SOULIER, P. (2000). Wavelet estimator of long-range dependent processes. *Stat. Inference Stoch. Process.* **3** 85–99. 19th “Rencontres Franco-Belges de Statisticiens” (Marseille, 1998).

BAYRAKTAR, E., POOR, H. V. and SIRCAR, K. R. (2004). Estimating the fractal dimension of the S&P 500 index using wavelet analysis. *Int. J. Theor. Appl. Finance* **7** 615–643.

COHEN, A. (2003). *Numerical analysis of wavelet methods*, vol. 32 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam.

CRAIGMILE, P. and PERCIVAL, D. (2005). Asymptotic decorrelation of between-scale wavelet coefficients. *IEEE Trans. Inform. Theory* **51** 1039–1048.

CRAIGMILE, P., PERCIVAL, D. and GUTTORP, P. (2005). Wavelet-based parameter estimation for polynomial trend contaminated fractionally differenced processes. To appear, *IEEE Trans. Signal Process*.

DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766.

DAUBECHIES, I. (1992). *Ten lectures on wavelets*, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GENÇAY, R., SELÇUK, F. and WHITCHER, B. (2002). *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press Inc., San Diego, CA.

GIRAITIS, L., ROBINSON, P. and SAMAROV, A. (1997). Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal.* **18** 49–61.

HURVICH, C. M., MOULINES, E. and SOULIER, P. (2002). The FEXP estimator for potentially non-stationary linear time series. *Stoch. Proc. App.* **97** 307–340.

HURVICH, C. M. and RAY, B. K. (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. Time Ser. Anal.* **16** 17–41.

MASRY, E. (1993). The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. *IEEE Trans. Inform. Theory* **39** 260–264.

MOULINES, E., ROUEFF, F. and TAQQU, M. (2005). A wavelet Whittle estimator of the memory parameter of a non-stationary Gaussian time series. Tech. rep., Ecole Nationale Supérieure des Télécommunications et Boston University.

MOULINES, E. and SOULIER, P. (2002). Long-range dependence: Theory and applications. In *Theory and applications of long-range dependence* (P. Doukhan, G. Oppenheim and M. Taqqu, eds.). Birkhäuser, Boston.

URL <http://math.bu.edu/people/murad/flyer-txt.pdf>

PAPANICOLAOU, G. C. and SØLNA, K. (2003). Wavelet based estimation of local Kolmogorov turbulence. In *Theory and applications of long-range dependence* (P. Doukhan, G. Oppenheim and M. Taqqu, eds.). Birkhäuser Boston, Boston, MA, 473–505.

PERCIVAL, D. B. and WALDEN, A. T. (2000). *Wavelet methods for time series analysis*, vol. 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

ROBINSON, P. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661.

SAMORODNITSKY, G. and TAQQU, M. S. (1994). *Stable non-Gaussian processes: stochastic models with infinite variance*. Chapman and Hall.

VEITCH, D. and ABRY, P. (1999). A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Trans. Inform. Theory* **45** 878–897.

VELASCO, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *J. Time Ser. Anal.* **20** 87–127.

VELASCO, C. and ROBINSON, P. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *J. Am. Statist. Assoc.* **95** 1229–1243.

YAGLOM, A. M. (1958). Correlation theory of processes with random stationary n th increments. *Amer. Math. Soc. Transl. (2)* **8** 87–141.

GET/TÉLÉCOM PARIS, CNRS LTCI, 46, RUE BARRAULT, 75634 PARIS CÉDEX 13, FRANCE.

E-mail address: moulines@tsi.enst.fr

E-mail address: roueff@tsi.enst.fr

DEPARTMENT OF MATHEMATICS AND STATISTICS, BOSTON UNIVERSITY BOSTON, MA 02215, USA.

E-mail address: murad@math.bu.edu